Manuscript: Ref. No. 2083

Title: Modeling the removal of Sunfix Red S3B from aqueous solution by Electrocoagulation process using artificial neural network

HA BUI MANH

Dear reviewers: We are very grateful for your comments to our manuscript. We revised the manuscript in accordance with your advice, and carefully proof-read the manuscript to minimize typographical, grammatical, and bibliographical errors. Here below is one-by-one response to your comments.

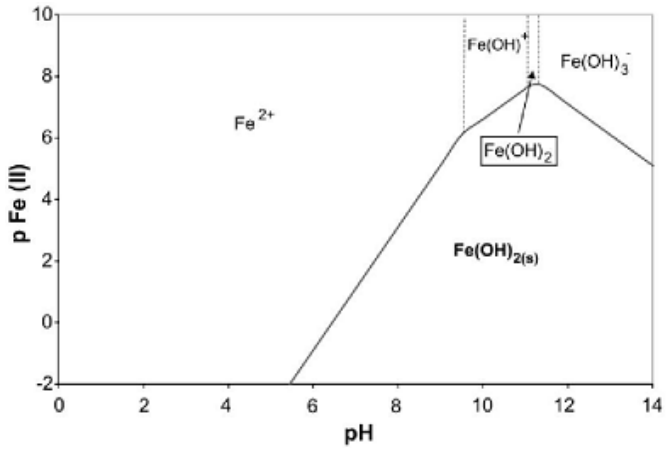
Referees' comments:

Reviewer# B:

1) Figure 1 proposed three mechanisms for dye removal in EC, precipitation, floatation and adsorption. Which mechanisms did the key role in the EC process, since it would be useful information to design the ANNs.

➔ Theoretically, effect of EC for removal pollutant in textile wastewater is the precipitation by either complexation or electrostatic attraction followed where iron ion play role as coagulants. To date, scientists still confused to find the key mechanism[1,2]. However, they supposed that the overall reaction principle is a combination of three removal mechanisms functioning synergistically. The dynamic process as the reaction progresses, and will certainly shift with changes in treatment conditions, operating parameters and in particular pollutant types.[3,4]

For example, variation in pH (see Figure below): at low pH value, the iron ion remains in the aqueous phase are Fe^{n+} (n=2 or 3). Therefore, in this case, precipitation should be predominant mechanism because dye could react directly with Fe^{n+} ion. On the contrary, at high pH value iron exits dominantly in form of $Fe(OH)_{n(s)}$. So, adsorption should be the predominant mechanism. Such phenomena are complicated and should be modelled by software model like ANN[5].



Predominance zone diagram for Fe(II) chemical species in aqueous solution [6]

REFERENCES

1. H. A. Moreno-Casillas, D. L. Cocke, J. A. G. Gomes, P. Morkovsky, J. R. Parga, E. Peterson, *Sep. Purif. Technol.* **56** (2007) 204.
2. V. M. Monsalvo, *Ecological Technologies for Industrial Wastewater Management: Petrochemicals, Metals, Semi-Conductors, and Paper Industries*, Apple Academic Press, Ontario, Canada, 2016, p. 202-205.
3. P. K. Holt, G. W. Barton, C. A. Mitchell, *Chemosphere* **59** (2005) 355.
4. M. Y. A. Mollah, R. Schennach, J. R. Parga, D. L. Cocke, *J. Hazard. Mater.* **84** (2001) 29.
5. S. M. Mirsoleimani-azizi, A. A. Amooey, S. Ghasemi, S. Salkhordeh-panbechouleh, *Industrial & Engineering Chemistry Research* **54** (2015) 9844.
6. C. Barrera-Díaz, M. Palomar-Pardavé, M. Romero-Romo, S. Martínez, *J. Appl. Electrochem.* **33** (2003) 61.

2) On Page 3, line 60, Fig. S-1 was missing.

➔ We have already uploaded the figure as a supplementary file. Please check.

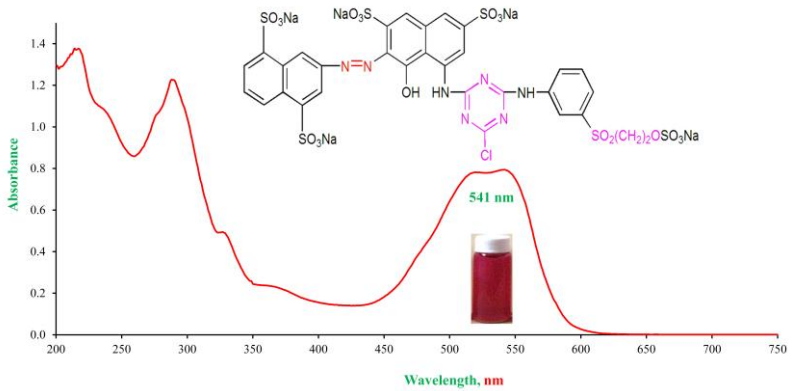
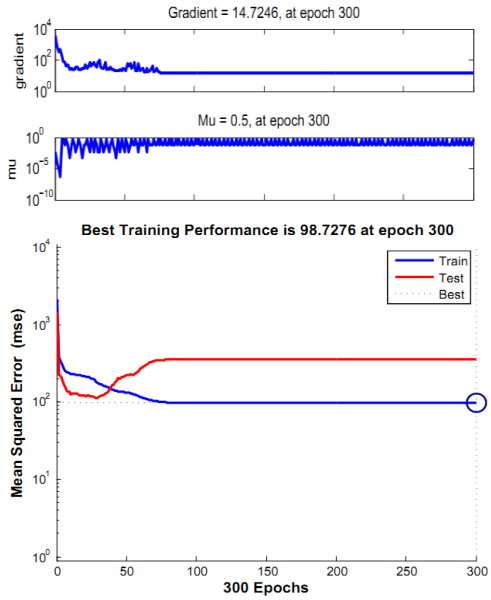
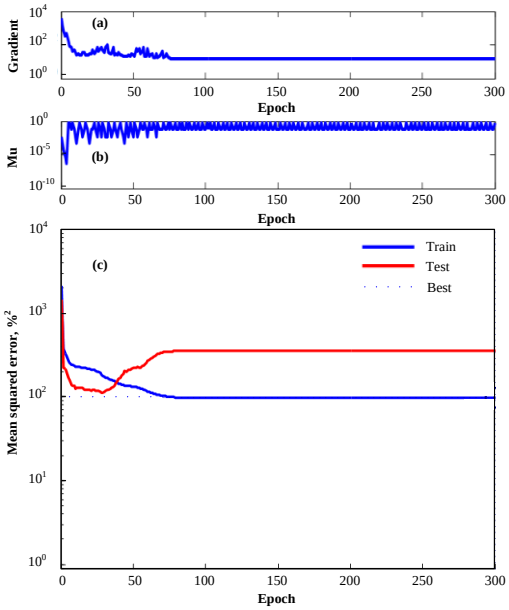


Fig. S-1. Chemical structure and UV-Visible spectral properties of Sunfix Red S3B

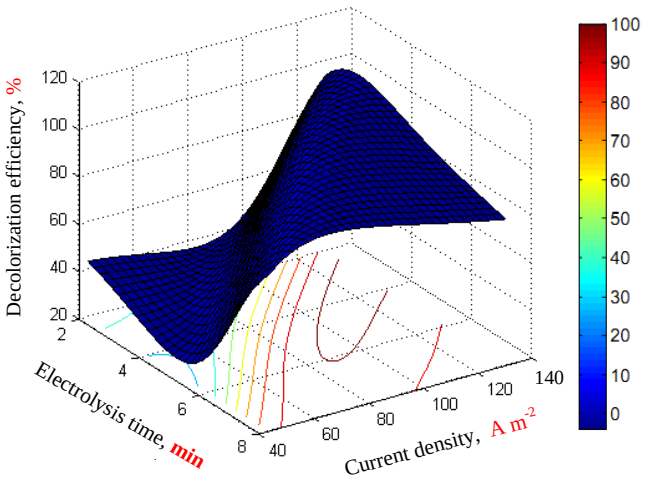
3) On Page 8, the quality of Fig. 6 was not good, specifically Y-axis needs high resolution. In addition, Fig.6 contains three sub-figures and need more information on captions and discussion.

➔ We have already made the figure clearer and given more information on the figure caption and the text. However, the figure was provided by Matlab ANN toolbox so we couldn't change the Y-axis resolution (see table below).

Original file	Edited file
	

4) Based on the results shown in Table 1 and Figure 8, for decolorization, the first three most influential factors are current density (I_3) > electrolysis time (I_4) > initial pH (I_2), while the other two factors initial dye concentration (I_1) and sulphate concentration (I_5) have minor effect. However, Fig.9 investigated the response and contour plot of electrolysis time vs. sulphate concentration on the removal of COD. I am wondering why the author didn't investigate the response and contour plot of the three most influential factors. Please explain.

➔ We only use response and contour plot to investigate the complicated case, i.e. when the COD removal efficiency of electrolysis time (I_4) and sulphate concentration (I_5) validated by two algorithms (Garson and Connection weight) contrast each other. In other cases, the two algorithms strongly agree about the effects of the most influential factors, so we don't investigate further. For example, the response and contour plots of the I_3 and I_4 (the most influence factors on decolorization) are almost monotonic like the plot below.



Response and contour plot of electrolysis time vs. current density on the decolorization at IDC 90 mg L^{-1}, initial pH 5, sulphate concentration 500 mg L^{-1}

Reviewer # C:

1) The number of data points used for model development needs to be presented.

➔ We already added the number of data points as your suggestion in Artificial neural network section (on page 5).

2) Descriptive statistics of used variables also need to be presented.

➔ We have already added the descriptive statistics and histogram distribution for the variables in Table S-3 and Figure S2 please find that information.

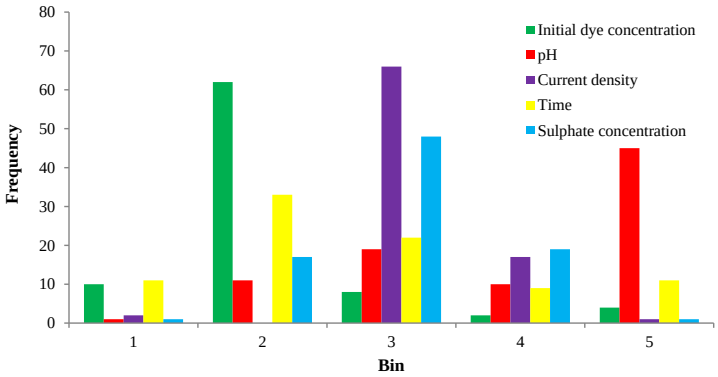


Fig. S-2. Distribution histogram of the variables

Table S-III. Descriptive statistics of variables (n=86) used in the construction of the ANN

Variable	Minimum	Maximum	Mean	Standard deviation	Variance
Initial dye concentration, mg L^{-1}	30	140	53.5	18.520	342.982
pH	3.13	12.04	8.6	2.624	6.887
Current density, A m^{-2}	43.3	130	81.1	18.245	332.873
Electrolysis time, min	2	8	4.7	1.252	1.568
Sulphate concentration, mg L^{-1}	500	2500	1395.3	382.840	146566.347

3) How the combination of exp. parameters (inputs) was determined? Usually, some experiential design is applied in order to "cover" entire regression surface.

➔ We have already added the experimental design section (on page 4), please check!

4) Some relative metrics (*e.g.* MAPE) need to be used for model evaluation.

➔ Thanks for your suggestion; we have already supplied MAPE, R^2 and RMSE value in Table S-4, please find that information.

Table S-IV. The effect of topology change on ANN's performance

Number of neuron in hidden layer	Total set			Training set		Test set	
	R^2	RMSE, %	MAPE, %	R^2	RMSE, %	R^2	RMSE, %
1	0.683	14.842	24.987	0.713	14.369	0.540	16.506
2	0.686	14.777	24.912	0.747	13.286	0.451	19.401
3	0.730	13.685	22.129	0.803	11.668	0.452	19.506
4	0.768	13.002	20.637	0.818	10.592	0.526	19.592
5	0.767	12.729	21.322	0.792	12.045	0.667	15.036
6	0.780	12.361	21.905	0.804	11.538	0.689	15.069
7	0.824	11.045	20.154	0.857	9.730	0.722	15.007
8	0.787	12.173	17.028	0.856	9.899	0.546	18.380
9	0.813	11.371	16.972	0.923	7.194	0.416	20.610
10	0.817	11.269	17.833	0.919	7.526	0.424	19.817
11	0.807	11.569	17.473	0.896	8.353	0.406	19.389
12	0.800	11.798	17.662	0.901	8.393	0.375	19.973
13	0.807	11.591	17.869	0.901	8.427	0.372	19.329
14	0.784	12.262	17.501	0.883	9.091	0.362	20.155
15	0.780	12.360	16.981	0.910	7.734	0.370	22.449
16	0.798	11.843	17.720	0.926	7.122	0.326	21.874
17	0.808	10.803	16.638	0.917	7.613	0.349	20.486

5) "With less than 5 hidden neurons, we observed the sign of underfitting where coefficients of determination of both training set and testing set is low. On the other

hand, with more than 7 hidden neurons, the sign of overfitting were observed that coefficient of determination of training set is high while that of testing set is low, i.e. the network lacks capability to predict untrained data". Test set must be "unseen" during the model development, and therefore it should not be used for the adjustment of parameters. This is a fundamental error. Determination of H neurons needs to be redone using *e.g.* 20% of training dataset for validation or by cross-validation.

➔ About validation, we use Bayesian regulation algorithm which does not require a validation set (see on page 8). "Validation set" here is meant to be used to determine a stopping point for the back-propagation algorithm, not to find the optimal number of hidden neurons. Hence, our "test set" is used in place of the "validation test" (in the second sense) to find the optimal number of hidden neurons. Then about the final "test set" for evaluating the final model, we actually have not prepared it!

6) Fig 5. and 6 are redundant.

➔ We have put the Figure 5 in supplemental section and revised the Figure 6 for the sake of clarity.

7) RMSE is the absolute indicator and, therefore, it has a unit!

➔ In this case, the unit of RMSE are % and it indicated in Figure 4. We have added the unit for all the manuscript to make it clear.

8) In Fig 7. the predicted removal is >100% for some data points?

➔ Yes, they are values predicted by the model. Because the model cannot predict exact values (must have errors) and the actual values are so close 100% (e.g decolorization efficiency got 99.5% at IDC 50 mg L^{-1}, pH 11, current density 108 A m^{-2}, electrolysis time 6 min, and sulphate concentration 2000 mg L^{-1}, respectively), it cannot help predicting out-of-bound values (>100%).

9) Comparison with similar studies is missing.

➔ We have already added several similar studies to improve the quality and clarity of the manuscript.

Beside revised the manuscript as your suggestions, we also adjusted some sentences in Introduction section, corrected table I and Figure 8 (unit) to improve the presentation.

We think those are excellence comments in for this manuscript and we are so glad to finish.