Internship report

Yang Shen, School of Aerospace Engineering, Tsinghua University

Komiya Lab, Institute of Fluid Science, Tohoku University

Abstract

This study focuses on the diffusion phenomena in ethanol-water systems. The diffusion coefficient of ethanol-water systems has a strong concentration dependency originating from intermolecular interactions and the association of molecules in solution. To quantitatively evaluate the concentration dependency in the diffusion process, the inverse analysis was conducted by genetic algorithm with deterministic crowding, finite differential method and weighting method.

Keywords

diffusion coefficient, ethanol, concentration dependency, genetic algorithm, deterministic crowding, Inverse analysis

I. BASIC WORKFLOW

In this study, the diffusion process was assumed to be isothermal and one-dimensional free diffusion. Thus, the governing equation of this diffusion process can be approximate as:

$$\partial C/\partial t = \frac{\partial}{\partial x}\left(D(C)\frac{\partial C}{\partial x}\right) \tag{1.}$$

$$\mathrm{D(C)} = \mathrm{D_0} + \alpha \mathrm{C} + \beta \mathrm{C}^2 \tag{2.}$$

Where x [m] is location, t [sec] is time, C[-] is the standardized concentration, and D [m2/s] is the mass diffusion coefficient which is a function of concentration. $\mathrm{D_0}$, α and β are coefficients used to represent concentration dependence of the mass diffusion coefficients quantitively. In the numerical simulation, Eq. (1) was discretized using the finite differential method as follows:

$$\mathrm{C_i^{n+1}} = C_i^n + \frac{\mathrm{D(C)}\,\Delta t}{\Delta x^2}\left(C_{i+1}^n - 2C_i^n + C_{i-1}^n\right) \tag{3}$$

where the subscript i represents the i^{th} grid position, and the superscript n represents n^{th} time interval. The whole spatial length is 20mm, the number of the spatial grids is 250 and the time step is 2s. For the initial conditions, a concentration distribution was obtained from the visualization experiment. At arbitrary elapsed diffusion time, the difference between the experimental and numerical concentration distribution was compared using an objective function *f* for inverse analysis, which is given by

$$\mathrm{f(D)} = \frac{1}{1 + \sum_{\mathrm{i}=1}^{250}\left(c_{cal} - c_{exp}\right)^2} \tag{4}$$

Where 250 is the number of grids, c_{cal} and c_{exp} are the calculation concentration and experimental concentration of every different grid, respectively. In a typical experiment, there would be concentration data of thousands of pixels, to match the calculation results and weaken the stochastic error form data, the experimental concentration was interpolated into 250 points, which share the same position of spatial grids in FDM.

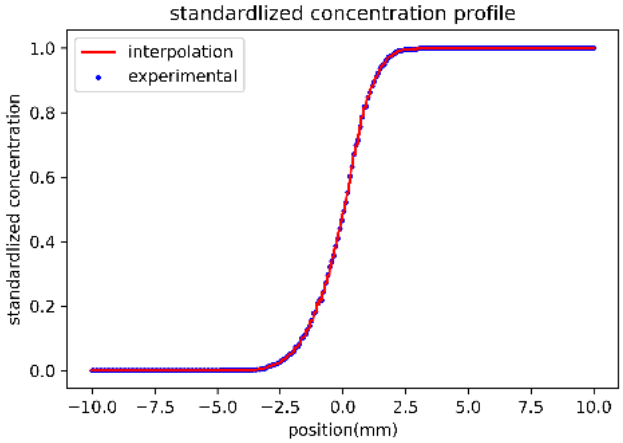

Fig. 1. Interpolation of experimental data

By maximizing the objective function *f(D)*, the integrated difference of concentration between experimental and numerical results in the target concentration range is minimized, and the diffusion coefficient and its concentration dependency are determined.

II. GENETICE ALGORITHM WITH DETERMINISTC CROWDING

Genetic algorithms are stochastic search and optimization methods imitating the process of evolution wherein a set of solutions evolves over a sequence of generations. In each generation, the fitness of each solution is evaluated, good solutions with a higher fitness are selected and poor solutions are eliminated. The selected solutions then undergo *crossover* and *mutation* process, which will create new solutions on the basis of good solutions in order to explore new solution space. After a sequence of generations, the global maximum of objective function is located and the corresponding solution is determined.

However, there are some limitations of traditional simple GA(SGA), one of the most significant demerits is premature convergence, which means the algorithm is possible to trap into local optimum when dealing with the high-dimension function with too many local optima. To solve this problem, many improvements were proposed, by comparing the performance of many different methods in our problem, we finally employ deterministic crowding (DC) variation in SGA, which is first proposed by De Jong[1] to preserve diversity in the population and to prevent premature convergence to local optima.

DC works as follows:

1) Initialization: randomly generate n individuals

2) Selection:

(REPEAT for g generations)

DO n/2 times:

1. select 2 parents, p1 and p2 randomly
2. cross and mutation, yielding c1 and c2
3. IF [d (p1,c1) + d (p2,c2)] ≤ [d (p1,c2) + d (p2,c1)]
   - IF f (c1) > f (p1) replace p1 with c1
   - IF f (c2) > f (p2) replace p2 with c1

   ELSE
   - IF f (c2) > f (p1) replace p1 with c2
   - IF f (c1) > f (p2) replace p2 with c1

Where n is the size of the population and g is the number of generations. d (p1,c1) is the distance between p1 and c1, which has different definitions like Euclidean distance and Hamming distance. For simplicity, here in our code we just defined it as the Euclidean distance between two phenotypes. For example, if p1 =$(D_{01}, \alpha_1, \beta_1)$ and c1 =$(D_{02}, \alpha_2, \beta_2)$, then

$$d(p1, c1) = \sqrt{(D_{01} - D_{02})^2 + (\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2} \quad (5)$$

Also, to enhance the precision of the algorithm, multi-point crossover and grey code were also applied. The basic parameters set in algorithm are shown as follows:

TABLE I. BASIC PARAMETERS OF ALGORITHM

Population size	Number of generations	Probability	
		crossover	mutation
50	100	0.82	0.07

The population size and number of generations were set by some trails, in this inverse analysis, the algorithm can converge with this setting. The probability of crossover and mutation was optimized by TPE approach (Tree-structured Parzen Estimator). For the detailed information please find in reference.[2]

III. VERIFICATION OF INVERSE ANALYSIS

A. *Verification of Finite Differential Method*

To verify the feasibility of the discretization scheme, the same simulation was conducted by FDM and COMSOL respectively. The initial condition taken is c = 8.3e-3 [mol/m3]from -10 to 0 [mm], c = 0 from 0 to 10 [mm], and the diffusion coefficient is set as 1e-9 [m2/s]. The calculation results are shown in figure 2.

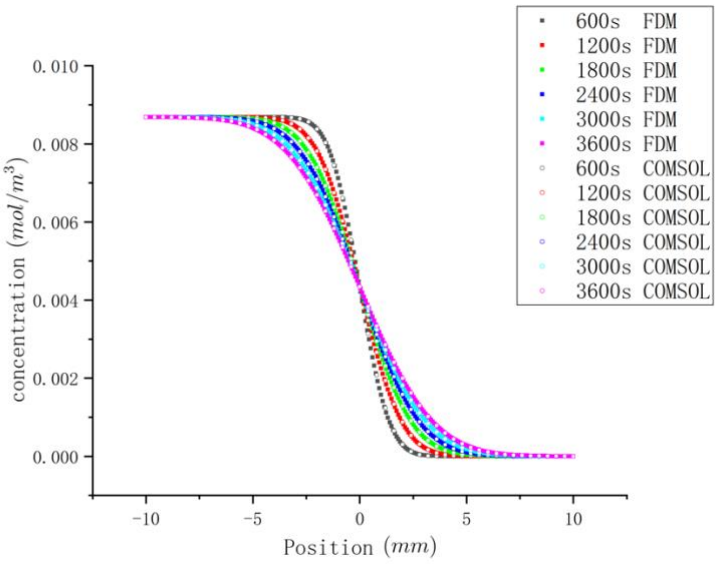

Fig. 2. The concentration filed calculated by FDM and COMSOL

It can be seen that the concentration profile calculated by FDM agrees very well with the one calculated by COMSOL (finite element method). Although numerical result is not completely identical to the experiment, for 1D diffusion model we can assume that the discretized scheme we take is reliable.

B. Verification of Algorithm

To validate the capability of the algorithm for inverse analysis, the test was conducted as follows:

The values of D_0, α and β were manually set at first, and the concentration profile calculated at different elapsed time by given coefficients were taken as 'experiment data'. Then examine whether the program can give the same values as manual set. The test results are shown in table 2.

TABLE II. TEST INVERSION RESULTS BY GENETIC ALGORITHM

METHOD	CODE	PARAMETER	VALUE	Predicted Value at Different Diffusion Time														Relative Error	Absolute Error
				600s/FITNESS		1200s/FITNESS		1800s/FITNESS		2400s/FITNESS		3000s/FITNESS		3600s/FITNESS		average			
Simple GA(SGA)	BINARY	D0	1.086	1.12	99.2	1.1	99.72	1.078	99.25	1.078	99	1.05	77.13	1.133	94.92	1.093167		1%	0.007166667
		α	-0.86	-1.03		-1		-0.832		-0.89		-0.965		-1.1		-0.9695		13%	0.1095
		β	0.57	0.724		0.712		0.554		0.605		0.712		0.787		0.682333		20%	0.112333333
	GREY	D0	1.086	1.13	96.25	1.13	87.09	1.11	97.96	1.086	93.24	1.125	95.41	1.11	94.6	1.115167		3%	0.029166667
		α	-0.86	-1.08		-0.927		-0.91		-1.06		-1.016		-1.05		-1.00717		17%	0.147166667
		β	0.57	0.787		0.58		0.611		-0.8		0.712		0.737		0.437833		23%	0.132166667
Deterministic Crowding GA(DCGA)	BINARY	D0	1.086	1.11	99.92	1.11	99.83	1	99.41	1.1	99.42	1.078	99.15	1.09	99.8	1.081333		0%	0.004666667
		α	-0.86	-0.997		-0.959		-0.92		-1.07		-0.8		-0.83		-0.92933		8%	0.069333333
		β	0.57	0.7		0.66		0.62		0.77		0.517		0.54		0.6345		11%	0.0645
	GREY	D0	1.086	1.09	99.92	1.1	99.86	1.09	99.97	1.09	99.85	1.09	99.58	1.09	99.9	1.091667		1%	0.005666667
		α	-0.86	-0.85		-0.94		-0.87		-0.851		-0.88		-0.84		-0.87183		1%	0.011833333
		β	0.57	0.554		0.64		0.58		0.554		0.586		0.554		0.578		1%	0.008

From the results, there is a big discrepancy between the real values and the inverted results calculated by SGA, and we can see that the deterministic crowding technique and grey code can enhance the performance of the program dramatically - the relative error of every coefficient was less than 1% which confirms that DCGA is workable in this inverse problem.

IV. RESULTS

Concerning for real experimental data, the different time interval will have an impact on inverse analysis because two diffusion field with an short elapsed time may be more vulnerable to the random disturbance, different combinations of start-end

couples were chosen. For example, START = 1200s, END = 3000s means that the initial concentration profile chosen is the diffusion field in 1200s and the end is the one in 3000s. The inversion results are shown in figure 4.

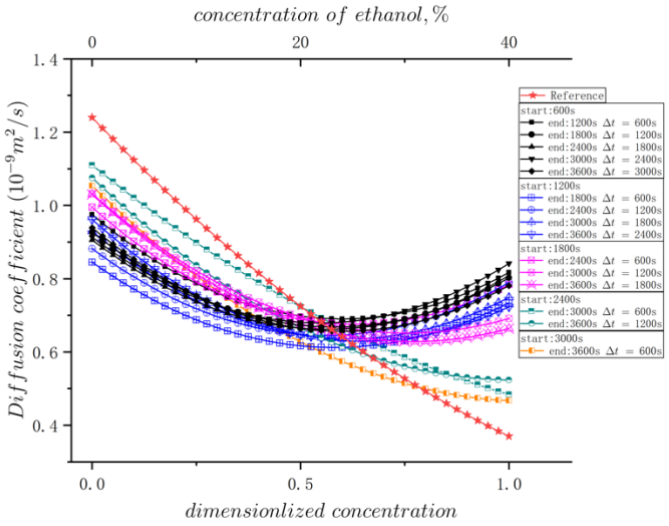

Fig. 4. Inversion results

The curve marked as *Reference* in the figure 4 was from Komiya's work[3], who collected the experimental data of diffusion coefficients in ethanol-water solution reported in many literatures and fitted them with quadratic function. From the results, it can be seen that when choosing early stages (before 2400s) as initial concentration profile, the concentration dependency of mass diffusion coefficient is strongly quadratic, which contradicts the relationship got from experiments, but after 2400s, the calculated dependency becomes linear which shows the very different tendency. To explain this phenomenon, the assumption we proposed is that before 2400s, the injection disturbance still had a strong impact on the diffusion process, till 2400s, it didn't dominate the system any more. To verify this hypothesis and quantify this injection disturbance, we propose a weighting method for inverse analysis which is shown in figure 5. We divide the whole diffusion region into 5 parts, f1 and f5 are constant concentration part, which is the least important for the inverse analysis, f2 and f4 are transition parts which have the highest second-order derivate, f2 represents the low concentration part and f4 represents the high concentration part. f3 represents the middle linear concentration part which has the highest first-order derivate and very low second-order derivative. The revised fitness function is defined as:

$$f(D) = \frac{1}{1 + \sum_{i=0}^{x1}(c_{cal} - c_{exp})^2 * f_1 + \sum_{i=x1}^{x2}(c_{cal} - c_{exp})^2 * f_2 + \sum_{i=x2}^{x3}(c_{cal} - c_{exp})^2 * f_3 + \sum_{i=x3}^{x4}(c_{cal} - c_{exp})^2 * f_4 + \sum_{i=189}^{250}(c_{cal} - c_{exp})^2 * f_5} \quad (5)$$

By putting a different weighting factor to the fitness function in different region, stability of inverse analysis and injection disturbance were investigated. The details are shown in figure 6,7,8 and 9.

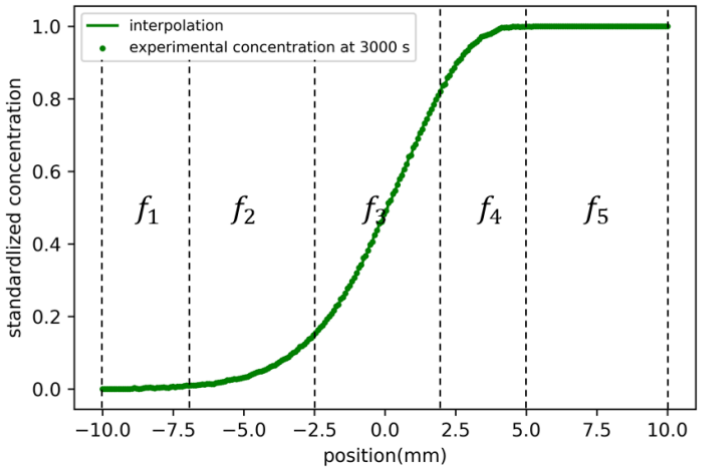

Fig. 5. Scheme for weighting method

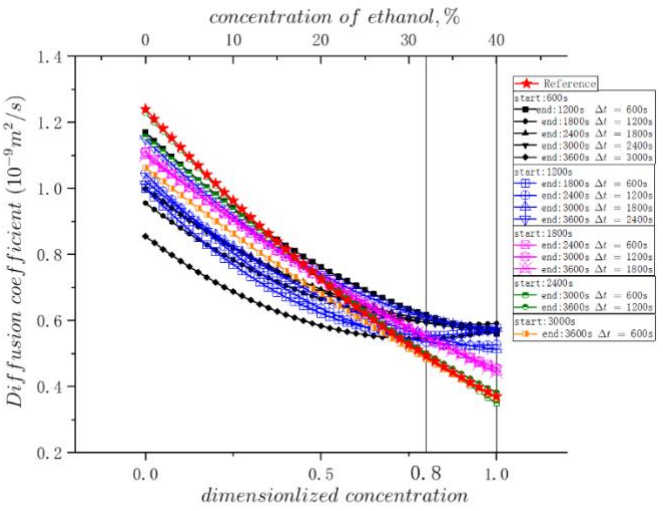

Fig. 6. High weight in high concentration region

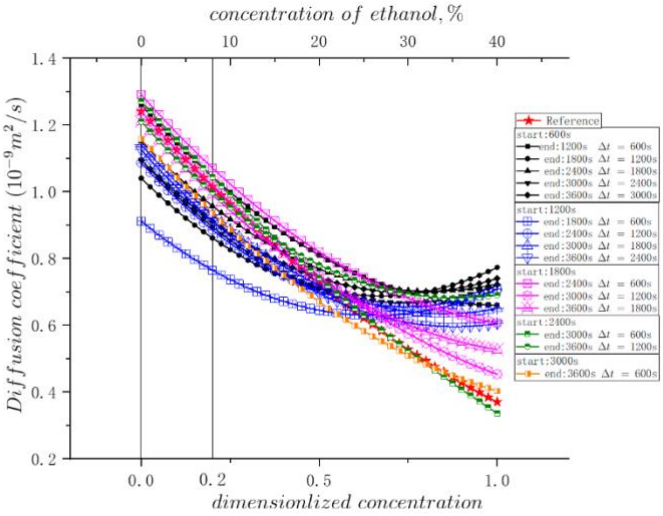

Fig. 7. High weight in low concentration region

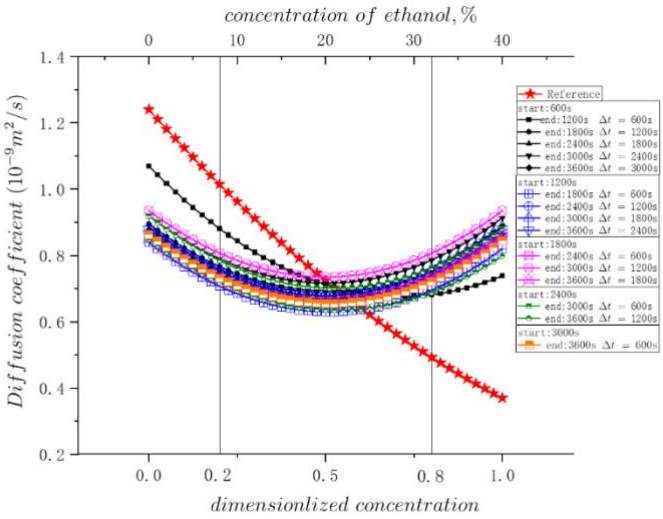

Fig. 8. High weight in middle concentration region

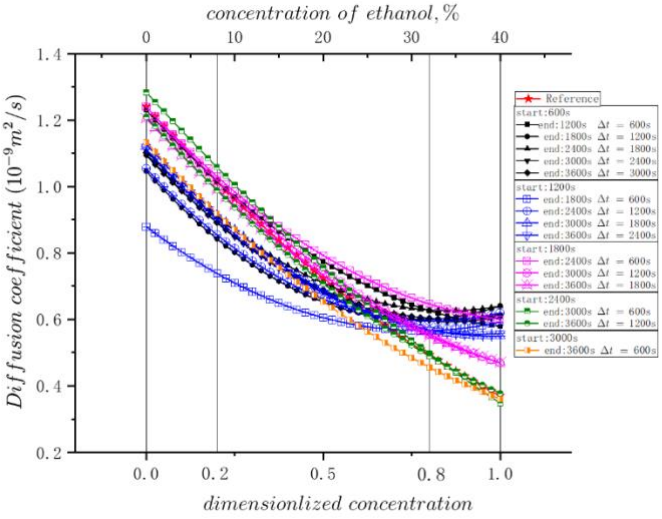

Fig. 9. High weight in high and low concentration region

with a high weighting in high concentration region (figure 6), it can be observed that when choosing a relatively early stage (before 2400s) as an initial concentration profile, there is a big discrepancy between calculation dependency and reference one. However, as start time increases, the difference becomes smaller. When the start time reaches 2400s, the difference nearly becomes zero in high concentration region. This result strongly suggests our previous hypothesis about injection disturbance, and it seems that only until 2400s the physical process becomes a pure diffusion process. Also, with a high weighting in low concentration region, it can be observed that since 1800s the calculation results converge to the literature data. It's also a valid argument to our assumption because in experiments, we injected ethanol to pure water, so compared with high concentration region, low concentration region is relatively less susceptible to injection. However, when putting a high weighting in middle linear part, all the dependencies GA predicted become strongly quadratic. From the results, we can dope out some characteristics of this inverse analysis: for the concentration transition parts, the solution space is more stable, although there may be small error in the experimental data, the dependency still gets around the real value. But for the middle linear part, it's very instable and chaotic, which means that even a little error in the experiment will lead to a big discrepancy to the result. By putting a high weighting in both low and high concentration region, we get the result shown in figure.9. it can be seen that for the combinations with $\Delta t = 600s$, the predicted dependency doesn't correspond with the ones calculated by the same initial concentration profile. It also fits our guess that this may be due to 600s interval for diffusion is too small, so that the concentration difference led by diffusion doesn't surpass the accumulative error in experiment and simulation a lot which contributes to the wrong result.

Finally, after the analysis of injection disturbance and the effect of time interval, the only combination valid in our experimental data is 2400s-3600s. And it can be seen this predicted dependency conforms to the literature dependency very well (figure.14) which also indirectly verifies that our hypothesis and analysis.

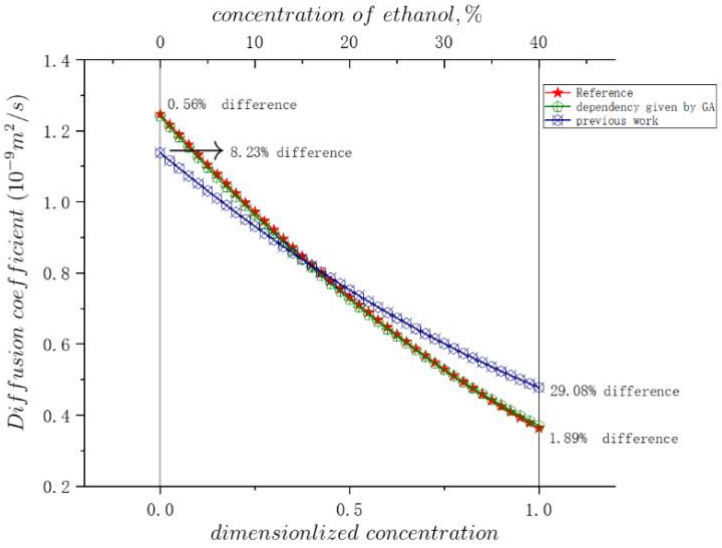

Fig. 10. High weight in high and low concentration region

V. CONCLUSIONS

In this study, the concentration dependency of diffusion coefficient in ethanol-water solution was determined by inverse analysis, the main conclusions of this study are summarized as following points:

1) Developed a program based on genetic algorithm with deterministic crowding, FDM and weighting method to determine the concentration dependency of mass diffusion coefficient and verified the feasibility of the algorithm

2) Analyzed the injection disturbance and examined the stability of the inverse analysis.

3) Acquired the concentration dependency of mass diffusion coefficient in ethanol-water solution in less 2% difference with the fitting of lots of experimental data reported in the literature

To get a more complete verification of this inverse analysis method, the performance of the program on more different kinds of solutions should be tested.

REFERENCES

[1] K. De Jong, "An analysis of the behavior of a class of genetic adaptive systems, 1975," *Univ. Michigan*. 1996.

[2] I. Resumen, I. I. Introducción, and I. I. I. R. D. E. Literatura, "Algorithms for Hyper-Parameter Optimization," pp. 1–9.

[3] A. KOMIYA, K. SATO, and J. OKAJIMA, "Measurement of concentration dependency of diffusion coefficient in ethanol-water solution under different storage condition," *J. Fluid Sci. Technol.*, vol. 13, no. 4, pp. JFST0030–JFST0030, 2018.