

Full-length papers

A Novel Multi-Dimensional Visualization Technique for Understanding the Design Parameters of Drug Formulations

Fumiyoshi Yamashita,^{1,*} Takayuki Itoh,³ Shuya Yoshida,¹ Mohammad K. Haidar,¹ Mitsuru Hashida.^{1,2}

¹Department of Drug Delivery Research, Graduate School of Pharmaceutical Sciences,
and ²Institute for Integrated Cell-Material Sciences, Sakyo-ku, Kyoto 606-8501, Japan.

³Department of Information Sciences, Faculty of Science, Ochanomizu University,
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan.

*To whom correspondence should be addressed:

Phone: 81-75-753-4535

Fax: 81-75-753-4575

E-mail: yama@pharm.kyoto-u.ac.jp

Abstract

The quality-by-design concept is a new regulatory paradigm for pharmaceutical development, while the response surface method (RSM) is a promising approach for understanding design parameters for drug formulations. RSM aims to provide a visual image to support statistical design and analysis of experiments. However, neither contour plots nor 3D surface plots that have commonly been used can completely visualize interactions between the parameters within the design space, due to their limited dimensionality. This article presents a visualization technique that can simultaneously display the responses to multi-dimensional factors by mapping N-dimensional data onto unique x-y coordinates, re-defined by recursive slice-and-dice subdivision of the 2D plane. The applicability of the technique was confirmed using published data on the design of nasal drug formulations.

Keywords: Design of experiments; Response surface method; Data visualization; Formulation design; Design space; Nasal drug formulations

1. Introduction

In light of the Quality Vision requirements set forth by the ICH, regulatory paradigms for pharmaceutical development are changing from quality-by-testing (QbT) to quality-by-design (QbD) (Food and Drug Administration CDER, 2004a). QbD is a systematic, risk-based approach to attaining desirable quality through careful evaluation of all attributes that influence product quality, from early development through the entire product lifecycle. The QbD approach not only reduces or eliminates compliance actions, penalties and recalls of products but also simplifies regulatory acceptance of post-approval manufacturing changes (Food and Drug Administration CDER, 2004b).

Design of experiments (DOE) is a structured and organized technique to determine the factors that may influence a product or process. DOE overcomes the problem associated with one-component-at-a-time experiments that cannot clarify interactive effects between factors. Response surface method (RSM) designs, including the central composite design and Box-Behnken design, are DOE methods that fit mathematical equations to experimental data to determine a set of design parameters that optimize a response. A mathematical model can provide an understanding of input-output relationships through a visual image of the response surface, such as contour plots or 3D surface plots. Recently, the importance of DOE and RSM technologies has been increasingly emphasized by advocates of the QbD concept in pharmaceutical development (Guo, 2008; Shah, 2007).

This article presents a novel visualization technique, named HyperDEC (Hyper-Dimensionally Embedded Cuboids), which helps us comprehensively

understand the effects of design parameters on the properties of drug formulations. HyperDEC overcomes the limitation of contour plots or 3D surface plots, which cannot simultaneously display the effects of more than two variables. HyperDEC maps N-dimensional data onto a 2D rectangular region defined by recursive slice-and-dice subdivision of the x - y plane. We have recently proposed data visualization techniques, called HeiankyoView, that display hierarchically structured data using colored icons and nested rectangular frames, and demonstrated their effectiveness in extracting information on the structure-activity relationship of aqueous solubility (Yamashita, 2006) and cytochrome P450-mediated drug metabolism (Yamashita, 2008). HeiankyoView and HyperDEC are similar in displaying hierarchically structured data but differ critically in spatial arrangement of icons. The effectiveness of HyperDEC was investigated by applying it to the data of Dayal et al. (2005) regarding the formulation of a nasal drug delivery system.

2. Theory of HyperDEC

HyperDEC maps N-dimensional data onto a 2D rectangular region defined by recursive slice-and-dice subdivision of the x - y plane. Figure 1 shows a 4-dimensional, 5-level plot, where x_3 - x_4 subgraphs are embedded into an x_1 - x_2 graph. Thus, 4-dimensional data, e.g., (0,0,2,2), (1,4,2,3), and (3,1,1,3), are placed in a unique position in the x - y absolute coordinate system.

Here, we derive a general solution for the x - y position, x -width (p_s), and y -height (q_s) of the data object that corresponds to the vector $(a_1, \dots, a_{2k-1}, a_{2k}, \dots, a_s)$

where $a_i = 0, 1, \dots, n-1$. In the recursive slice-and-dice procedure, odd-numbered discretizations are conducted vertically while even-numbered discretizations are done horizontally. As shown in Fig. 2, let us assume that the original position, x -width, and y -height before $(2k-1)^{\text{th}}$ discretization are (x_{2k-2}, y_{2k-2}) , p_{2k-2} , and q_{2k-2} , respectively and that the object was divided into n pieces with a space size of $\frac{p_{2k-2}m_{2k-1}}{n-1}$ for each.

When the $(2k-1)^{\text{th}}$ element of the vectorial data is a_{2k-1} , relationships between x_{2k-2} and x_{2k-1} and between y_{2k-2} and y_{2k-1} are:

$$x_{2k-1} = x_{2k-2} + a_{2k-1} \frac{p_{2k-2}}{n} \left(1 + \frac{m_{2k-1}}{n-1} \right)$$

$$y_{2k-1} = y_{2k-2}$$

Relationships in x -width and y -height are:

$$p_{2k-1} = \frac{p_{2k-2}}{n} (1 - m_{2k-1})$$

$$q_{2k-1} = q_{2k-2}$$

On the other hand, horizontal slicing at the $2k^{\text{th}}$ discretization gives:

$$x_{2k} = x_{2k-1}$$

$$y_{2k} = y_{2k-1} + a_{2k} \frac{q_{2k-2}}{n} \left(1 + \frac{m_{2k}}{n-1} \right)$$

$$p_{2k} = p_{2k-1}$$

$$q_{2k} = \frac{q_{2k-2}}{n} (1 - m_{2k})$$

When the original position, x -width, and y -height are (x_0, y_0) , p_0 , and q_0 , respectively, these recursive formulae give:

if $s (= 2t - 1)$ is an odd number,

$$\text{position: } \left(x_0 + \sum_{k=1}^t \frac{a_{2k-1} p_{2k-2}}{n} \left(1 + \frac{m_{2k-1}}{n-1} \right), y_0 + \sum_{k=1}^{t-1} \frac{a_{2k} q_{2k-2}}{n} \left(1 + \frac{m_{2k}}{n-1} \right) \right)$$

$$\text{x-width: } p_0 \prod_{k=1}^t \frac{1-m_{2k-1}}{n}$$

$$\text{y-height: } q_0 \prod_{k=1}^{t-1} \frac{1-m_{2k}}{n}$$

if $s (= 2t)$ is an even number,

$$\text{position: } \left(x_0 + \sum_{k=1}^t \frac{a_{2k-1} p_{2k-2}}{n} \left(1 + \frac{m_{2k-1}}{n-1} \right), y_0 + \sum_{k=1}^t \frac{a_{2k} q_{2k-2}}{n} \left(1 + \frac{m_{2k}}{n-1} \right) \right)$$

$$\text{x-width: } p_0 \prod_{k=1}^t \frac{1-m_{2k-1}}{n}$$

$$\text{y-height: } q_0 \prod_{k=1}^t \frac{1-m_{2k}}{n}$$

Gradual reduction of the spacings, *i.e.*, m_i , for every dimension enables us to discriminate between nested hierarchical data objects. The technique is an extension of the dimensional stacking technique (LeBlanc, 1990) that involves discretization and recursive embedding of a pair of dimensions. The unique features of HyperDEC include filtering of interesting factors or data and simultaneous representation of two response variables with the heat color and z-height of cuboidal graphic objects.

The software product was developed in the Java programming language (Java SE 6) with Java bindings for OpenGL (JOGL) API. In addition to zooming, panning and rotating, functions that allow the selection of data objects on a graphical user interface were implemented in the software.

3. Features of HyperDEC

Figure 3 shows a typical example of HyperDEC images, together with the conventional contour plots that are widely used for understanding input-output relationships in DOE studies. As a simple case, assume here that responses are displayed as a function of four variables. To create a contour plot, two explanatory variables are selected as variables of interest, while other explanatory variables are held constant. When the constant variables are changed in a systematic, stepwise manner, responses with respect to all explanatory variables can be computed. Figure 3B is a systematic arrangement of all contour plots, which provides essentially the same information as the HyperDEC image shown in Fig. 3A, which consists of nested dimensions.

Visualization of all information does not always aid understanding of a relationship. As the dimensionality increases, the visual image becomes more complex, even if the dimensions are arranged in an orderly manner. Filtering of a graphics image is a useful technique to extract information about its contents by highlighting regions of interest. In the HyperDEC images, a set of data varying in one dimension only are aligned either horizontally or vertically at regular intervals. Figure 3C shows the images for the selection of each of two variables. Because this figure retains the same image structure as Fig. 3A, it is effective in discussing the effects of selected variables from a global perspective. In addition, plots having differing constant variables can be generated simply by changing the data objects that are highlighted. Contour plots such as those shown in Fig. 3D are easy to interpret, but present limited information, due to their lack

of flexibility. Thus, the graphic filtering technique of HyperDEC allows us to investigate the effect of each variable easily without altering the data structure, leading to an unbiased, systematic understanding of the characteristics of the response functions.

4. Application to a DOE study of nasal drug formulation

4.1. Data Source

Data were taken from the publication of Dayal et al. (2005). They investigated the influence of changes in physicochemical properties on nasal formulation performance using a 5-factor, 3-level Box-Behnken experimental design on the combined responses of viscosity, droplet size distribution and drug release. The factors investigated were the amounts of components in a model drug formulation comprised of hydroxyurea (HU; 0–4%), hydroxyethylcellulose (HEC; 0–4%), polyethylene oxide (PEO; 0–4%), sodium chloride (NaCl; 0–30%), and calcium chloride (CaCl_2 ; 0–30%). According to their Box-Behnken designs, the authors prepared a total of 44 formulation combinations and assessed their physicochemical properties. In the present analysis, quadratic regression equations approximating the data were used for graphic representation.

4.2. Visualization of the relationships between formulation variables and aerosol droplet size

Figure 4 shows a screen capture image of HyperDEC, where the effect of all five formulation factors on the droplet size of nasal aerosols is displayed. Both color and z-height of the bars indicate droplet size. The viewer control panel on the right allows

the user to zoom, move and rotate the image object and to filter out uninteresting elements. The focus on data objects can be changed through the graphic user interface. Such functions implemented in HyperDEC allow the acquisition of both global and local characteristics of the data structure simultaneously.

Graphic filtering was carried out to provide better understanding of the HyperDEC image (Fig. 5). Here, the levels of both NaCl and HU were set to be medians (15% and 2%, respectively): hence, it provided a graphic image with clusters of data objects that were located in the center of each of the corresponding inner subgraphs. The innermost dimension indicates CaCl_2 level, while the outermost left-right and bottom-top dimensions indicate HEC and PEO levels, respectively. When the image is scanned from bottom to top, the z-height of data objects becomes higher, indicating that PEO increases the droplet size, regardless of the levels of HEC and CaCl_2 . In contrast, the influence of HEC on the droplet size is not as simple. When the image is scanned from left to right, the z-height of data object clusters changes from a positive linear to a flat or gently parabolic pattern. This indicates that the effect of CaCl_2 on the droplet size becomes less significant with an increase in HEC: in other words, the effect of HEC on the droplet size is dependent on the CaCl_2 level.

Dayal et al. (2005), who published original data, mentioned that the addition of CaCl_2 to PEO shows a mild parabolic relationship in the droplet size of aerosols, with a reduction at low concentrations (10~15%) of CaCl_2 . As indicated in Fig. 5, however, PEO does not exhibit non-linear interaction with other ingredients. Moreover, the parabolic relationship that occurs with the addition of CaCl_2 is observed only in the presence of moderate or higher levels of HEC: in the absence of HEC, the addition of CaCl_2 exhibits a positive linear relationship with the droplet size. Comprehensive

re-evaluation using HyperDEC concludes that CaCl_2 interacts with HEC but not markedly with PEO.

4.3. Simultaneous visualization of effects on droplet size and solution viscosity

An additional feature of HyperDEC is the ability to represent two different response variables simultaneously using the color temperature and z-height of data objects (Fig. 6). In Fig. 6, the color temperature indicates the viscosity of the solution, while the z-height indicates aerosol droplet size. Each dimension of the image is the same as in Fig. 5. When the color temperature in Fig. 6A is considered, it is found that addition of HEC and PEO increases the viscosity of the solution but, as indicated by the color temperature in the innermost dimension, CaCl_2 has minimal affect on the solution viscosity. At the highest PEO level (the top of the outermost graph), NaCl (the bottom-top dimension of inner subgraphs) appears to reduce the solution viscosity. When both the color and z-height are compared, the relationship between the viscosity and droplet size can be explored. In general, as the PEO and HEC levels increase, both color temperature and z-height of data objects become higher (Fig. 6A). However, when both PEO and CaCl_2 are at their highest levels (4% and 30%, respectively), the color of the data objects becomes more red but the z-height becomes lower with an increase in HEC (Fig. 6B). Thus, under these conditions, HEC increases the solution viscosity but decreases the droplet size. It should be noted that smaller and more viscous droplets are better in terms of both spatial distribution and residence time in the nose. Thus, such a visual image would assist in the optimization of a formula for nasal spray formulations.

5. Discussion

Data visualization is an increasingly important area of technology research and development. One needs to manage massive amounts of data collected from various sources and optimize engineering and manufacturing processes. Data visualization techniques offer an intuitive way to acquire knowledge from complex data and transfer information to other people. An important issue for data visualization techniques is how to visualize data objects containing multidimensional information. One way to visualize multidimensional data in a two-dimensional (2D) or three-dimensional (3D) space is to reduce data dimensionality. Principal component analysis and partial least squares methods are widely used dimensionality reduction methods, where linear combinations of variables are explored systematically to explain the largest amount of variation in the original data set. Reduction of dimensionality by these methods is primarily based on collinearity of variables. However, these methods are ineffective in the present study because the formulation variables are independent of one another. Many multi-dimensional visualization techniques, such as Parallel Coordinates (Inselberg, 1985) and Star Glyphs (Andrew, 1972), are intended to visualize all attributes of data but not to identify multi-dimensional causal relationships. On the other hand, nested-graph methods, such as Worlds-within-Worlds (Feiner, 1990) and dimensional stacking (LeBlanc, 1990), may be useful for detecting multi-dimensional relationships. HyperDEC is an extension of dimensional stacking, the features of which include filtering of interesting factors and simultaneous representation of two response variables using cuboidal objects. In particular, as described in the preceding sections, the graphic

filtering functions implemented in HyperDEC allowed us to systematically investigate the effects of each explanatory variable on the response.

In common with conventional response surface methodologies, HyperDEC displays the results of simulations using a model function fitted to experimental data. In the present study, the application of HyperDEC was exemplified by a quadratic equation given by Dayal et al. (2005). In practice, any theoretical models, including artificial neural networks and spline functions, are compatible with HyperDEC, due to the input file for HyperDEC being simply an Extensive Markup Language (XML) format, where numerical simulation data are hierarchically structured. Moreover, adoption of the XML format makes HyperDEC flexible and extensible so that it can display various kinds of information in addition to response variables, e.g. confidence limits for individual data points. HyperDEC can also allot two different sets of data to both color and z-height of image data objects, further enhancing its capability.

In conclusion, we have developed a novel data visualization technique for the simultaneous display of a multi-factor response surface. Data visualization provides an intuitive understanding of the statistical evaluations obtained by the DOE and assists in the selection of important controllable parameters, i.e., design parameters from results of exploratory experiments. We believe HyperDEC will be a useful tool for identifying the design space for drug formulations under the QbD concept.

6. Acknowledgment

This research was supported in part by Grant-in-Aids for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- Andrews, D.F. (1972). Plots of high dimensional data. *Biometrics*, 28, 125.
- Dayal, P. Pillay, V., Babu, R.J., Singh, M. (2005). Box-Behnken experimental design in the development of a nasal drug delivery system of model drug hydroxyurea: characterization of viscosity, in vitro drug release, droplet size, and dynamic surface tension. *AAPS PharmSciTech*. 6, E573.
- Feiner, S., Beshers, C. (1990). Worlds within worlds: metaphors for exploring n-dimensional virtual worlds. *Proceedings of the 3rd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology*, ACM, New York, pp. 76-83.
- Food and Drug Administration CDER (2004a). Pharmaceutical cGMPs for the 21st century: A risk-based approach. http://www.fda.gov/cder/gmp/gmp2004/GMP_finalreport2004.htm.
- Food and Drug Administration CDER (2004b). Guidance for Industry: Q8 Pharmaceutical Development. <http://www.fda.gov/cder/guidance/6746f1.htm>.
- Guo, C., Stine, K.J., Kauffman, J.F., Doub, W.H. (2008). Assessment of the influence factors on in vitro testing of nasal sprays using Box-Behnken experimental design. *Eur. J. Pharm. Sci.* 35, 417.
- Inselberg, A. (1985). The plane with parallel coordinates. *Visual Comp.* 1, 69.
- LeBlanc, J., Ward, M.O., Wittels, N. (1990). Exploring N-dimensional databases. in: Kaufman, A. (Ed.), *IEEE Visualization: Proceedings of the 1st conference on Visualization '90*, IEEE Computer Society Press, Los Alamitos, pp. 230–237.

- Shah, R.B., Zidan, A.S., Funck, T., Tawakkul, M.A., Nguyenpho, A., Khan, M.A. (2007) Quality by design: characterization of self-nano-emulsified drug delivery systems (SNEDDs) using ultrasonic resonator technology. *Int. J. Pharm.* 341, 189.
- Yamashita, F., Itoh, T., Hara, H., Hashida, M. (2006). Visualization of large-scale aqueous solubility data using a novel hierarchical data visualization technique. *J. Chem. Inf. Modeling.* 46, 1054.
- Yamashita, F., Hara, H., Itoh, T., Hashida, M. (2008). Novel hierarchical classification and visualization method for multiobjective optimization of drug properties: application to structure-activity relationship analysis of cytochrome P450 metabolism. *J. Chem. Inf. Modeling.* 48, 364.

FIGURE CAPTIONS

Figure 1. An example of the HyperDEC plot, consisting of 4 dimensions (x_1 to x_4) and 5 levels (0 to 4 for each dimension). x_3 - x_4 subgraphs are embedded into an x_1 - x_2 graph.

Figure 2. Slice-and-dice method for mapping multi-dimensional data onto the x-y coordinate plane.

Figure 3. Comparison of a HyperDEC image with contour plots: A, a 4-dimensional, 5-level HyperDEC image; B, systematic arrangement of contour plots following 5x5 parameter scan simulation; C, filtered HyperDEC images for combinations of any two dimensions; D, contour plots for combinations of any two dimensions. In Figs. 3C and 3D, two constant variables are set at half the maximum values for each. Figures 3B and 3D were created using the *subplot* and *contour f* functions implemented in MATLAB 7.6 (The MathWorks, Inc., MA).

Figure 4. A screen shot of HyperDEC. Results of the design-of-experiments (DOE) study of a nasal spray formulation (Dayal et al., 2005) were represented as a HyperDEC image. Formulation variables comprise HEC, PEO, HU, NaCl and CaCl₂. Both the color temperature and z-height of data objects indicate the droplet size, while their x-y positions are determined by the levels of each formulation variable. Each dimension of these images and its scale are the same as in Fig. 5.

Figure 5. Graphic filtering of the HyperDEC image representing the droplet size of a nasal spray formulation. From the graphic image in Fig. 4, only data objects having the median levels of HU (2%) and NaCl (15%) are highlighted. The ranges of HEC, PEO, HU, NaCl and CaCl₂ for simulation were 0–4%, 0–4%, 0–4%, 0–30% and 0–30%, respectively.

Figure 6. HyperDEC images representing the solution viscosity (color) and droplet size (z-height) of nasal spray formulations: A, All data objects are represented; B, Data objects having the highest level of CaCl₂ in the absence or highest level of PEO are highlighted. Each dimension of these images and its scale are the same as in Fig. 5.

Figure 1

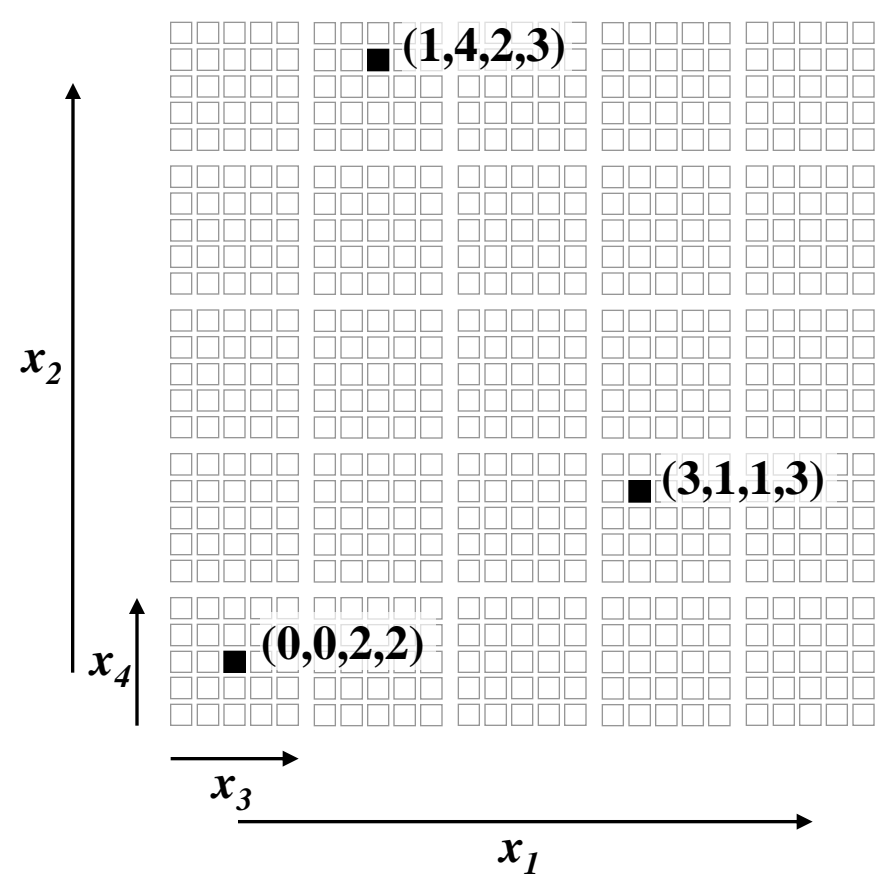
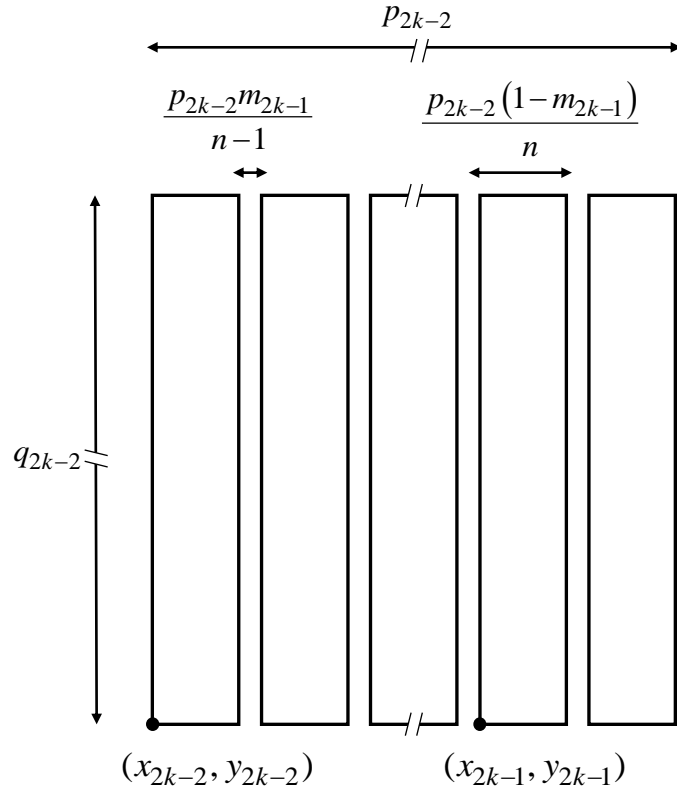


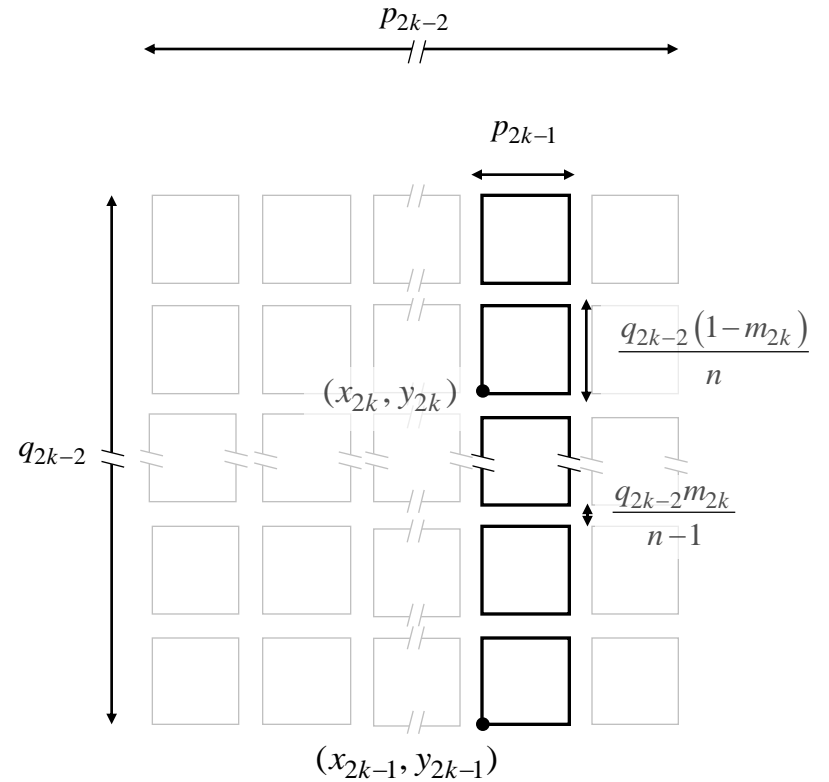
Figure 2

(A) $(2k-1)$ -th discretization



$$\begin{aligned} x_{2k-1} &= x_{2k-2} + a_{2k-1} \frac{p_{2k-2}}{n} \left(1 + \frac{m_{2k-1}}{n-1} \right) \\ y_{2k-1} &= y_{2k-2} \\ p_{2k-1} &= \frac{p_{2k-2}}{n} (1 - m_{2k-1}) \\ q_{2k-1} &= q_{2k-2} \end{aligned}$$

(B) $2k$ -th discretization



$$\begin{aligned} x_{2k} &= x_{2k-1} \\ y_{2k} &= y_{2k-1} + a_{2k} \frac{q_{2k-2}}{n} \left(1 + \frac{m_{2k}}{n-1} \right) \\ p_{2k} &= p_{2k-1} \\ q_{2k} &= \frac{q_{2k-2}}{n} (1 - m_{2k}) \end{aligned}$$

Figure 3

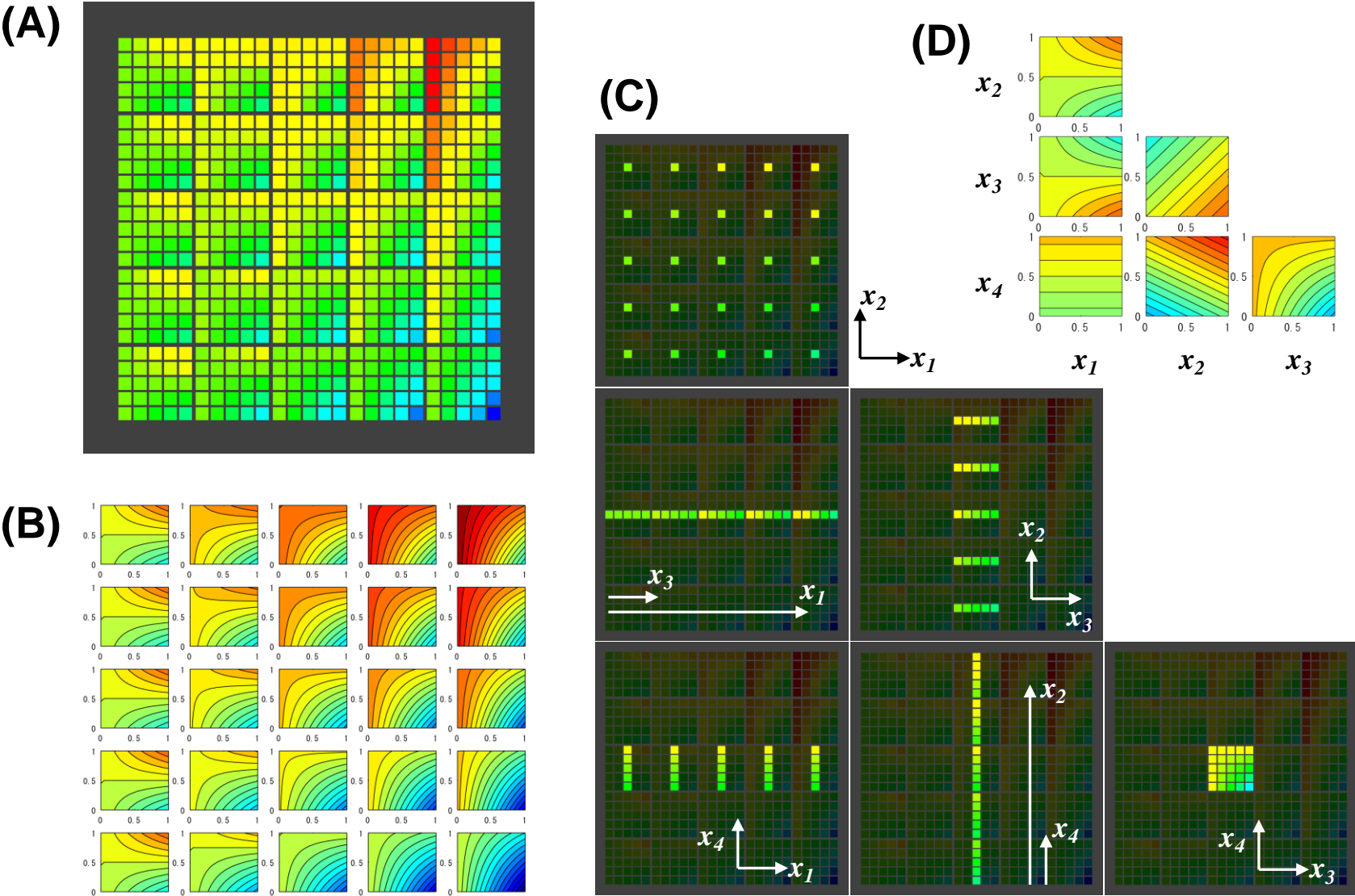


Figure 4

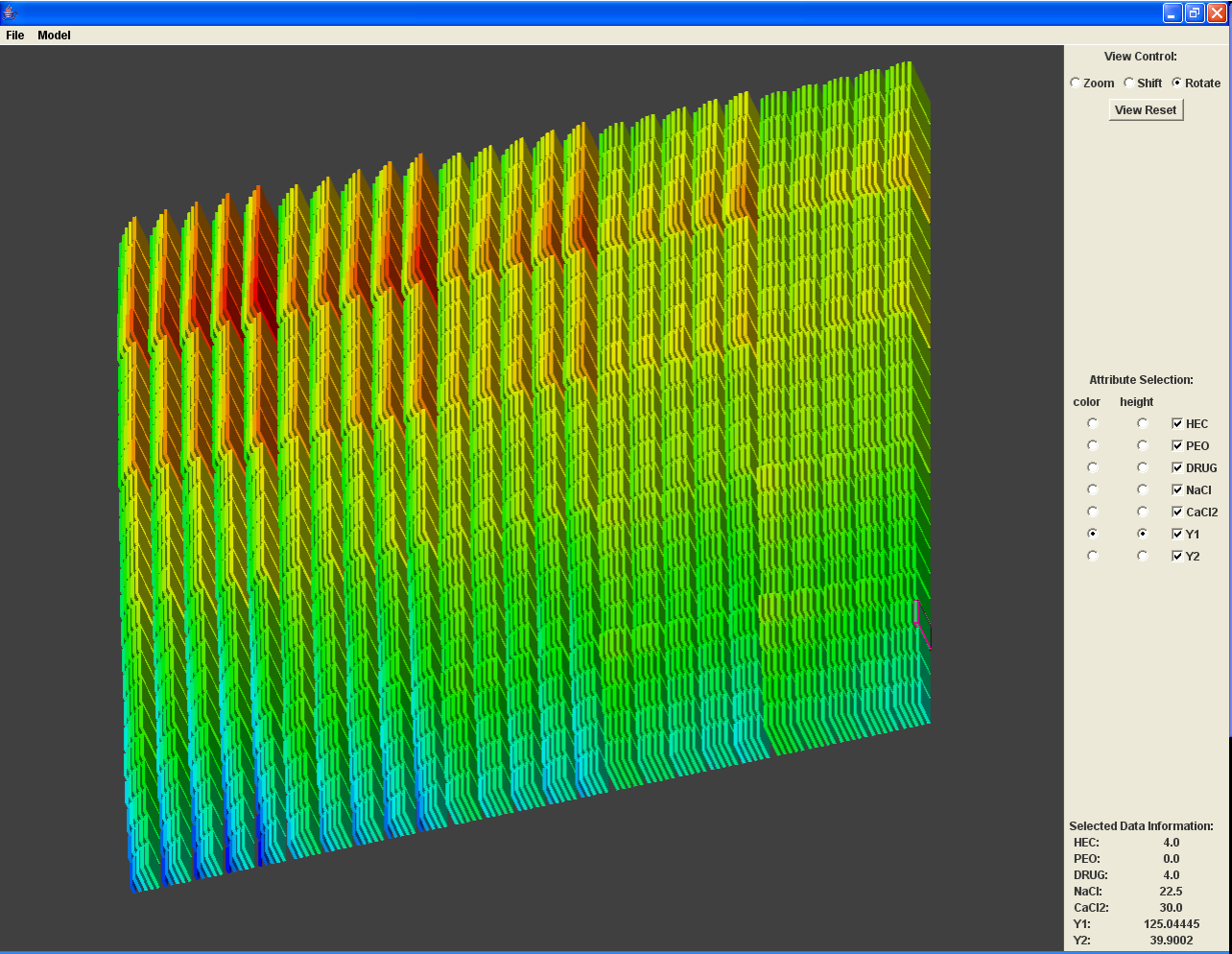


Figure 5

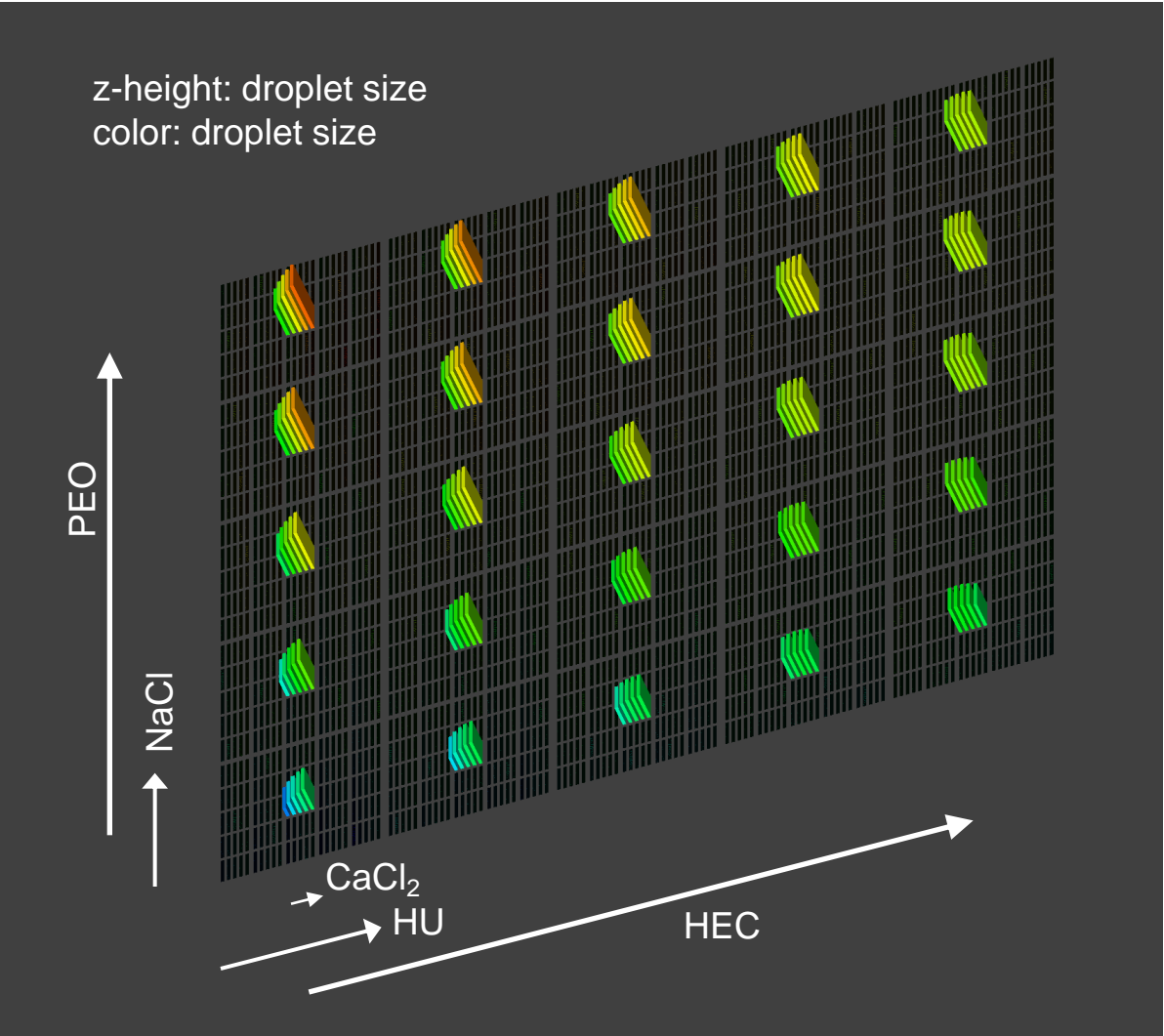
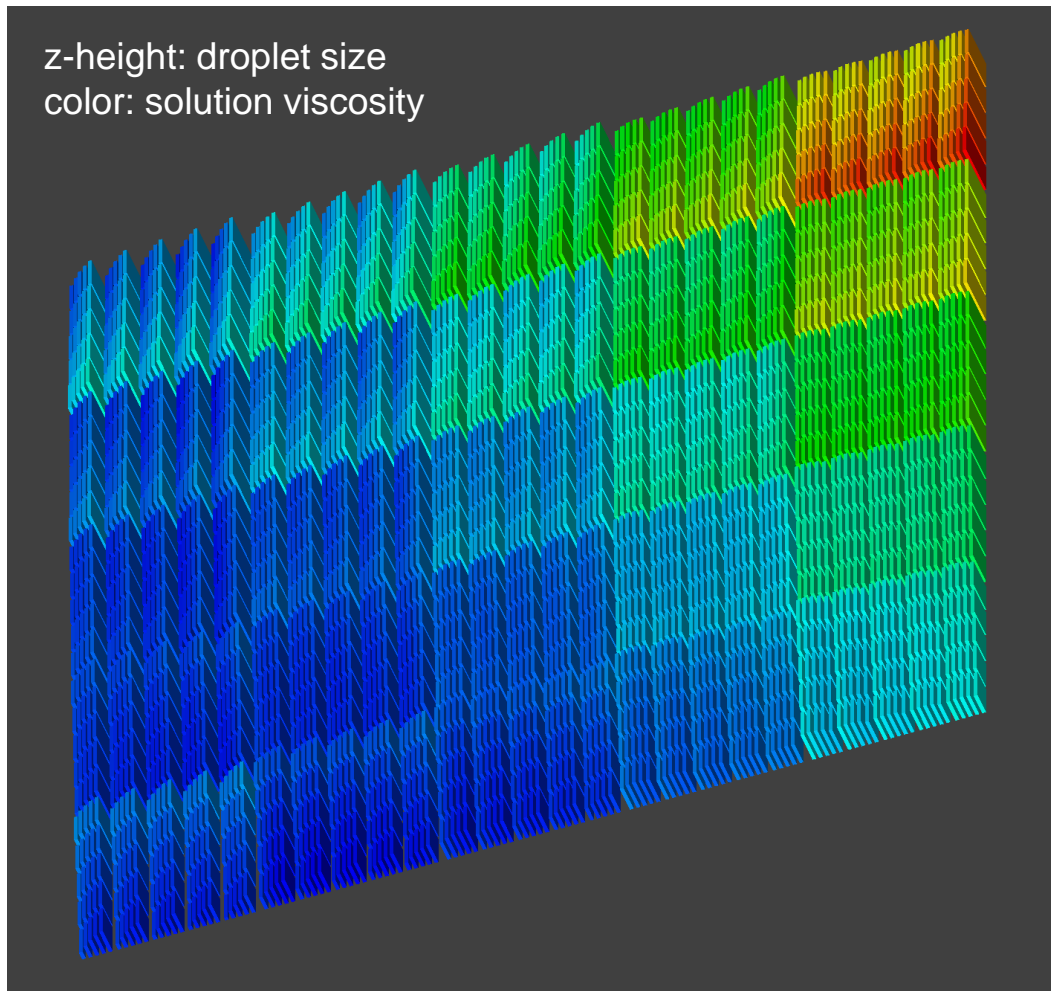


Figure 6

(A)

z-height: droplet size
color: solution viscosity



(B)

