ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
24 (2001) 343–352

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Visualization of fluid-bed granulation with self-organizing maps

Jukka T. Rantanen [a,*], Sampsa J. Laine [b], Osmo K. Antikainen [a], Jukka-Pekka Mannermaa [a], Olli E. Simula [b], Jouko K. Yliruusi [a]

[a] *Pharmaceutical Technology Division, Department of Pharmacy, University of Helsinki, Finland*
[b] *Neural Networks Research Centre, Laboratory of Computer and Information Science, Helsinki University of Technology, Finland*

Received 22 February 2000; accepted 20 May 2000

Abstract

The degree of the instrumentation of pharmaceutical unit operations has increased. This instrumentation provides information of the state of the process and can be used for both process control and research. However, on-line process data is usually multidimensional, and is difficult to study with traditional trends and scatter plots. The Self-Organizing Map (SOM) is a recognized tool for dimension reduction and process state monitoring. The basics of the SOM and the application to on-line data collected from a fluid-bed granulation process are presented. As a batch process, granulation traversed through a number of process states, which was visualized with SOM as a two-dimensional map. In addition, it is demonstrated how the differences between granulation batches can be studied. The results suggest that SOM together with new in-line process analytical solutions support the in-process control of the pharmaceutical unit operations. Further, a novel research tool for understanding the phenomena during processing is achieved.

Keywords: Granulation; NIR; Process monitoring; Reflectance; Self-organizing maps (SOM)

1. Introduction

The manufacture of pharmaceutical products usually requires a sequence of batch unit operations. These units are controlled by process operators who often collect information on the process with on-line measurement devices. An important part of their process-controlling skill is to determine the state of the process from these on-line measurements. A skilled process operator recognizes and evaluates process states and can advance the process from one state to another. Thus, the operator performs high-dimensional on-line data analysis. However, an inexperienced analyst or a novice process operator needs tools to obtain similar results. In this study, the self-organizing map (SOM) [1] was applied for this purpose.

The demonstration process, fluid-bed granulation (FBG), is a multivariate process with several

* Corresponding author. Tel.: +358-9-19159141; fax: +358-9-19159144.
E-mail address: jukka.rantanen@helsinki.fi (J.T. Rantanen).

PII: S0731-7085(00)00458-1

interacting parameters. FBG consists of three sequential phases: mixing, granulation (spraying), and drying. Each phase can be considered an individual process with unique process conditions.

Traditionally, the level of automation of pharmaceutical unit operations has been low. This has been due to the high documentation requirements of new process control practices. However, the development of process analytical chemistry [e.g. near-infrared (NIR) techniques] will improve the in-process control of pharmaceutical manufacture. NIR spectroscopy has been applied for measurement of moisture during wet granulation processes [2–6], homogeneity of mix [7,8], and measurement of different coating process parameters [9–11]. Han and Faulkner [12] developed a NIR method for different stages of tablet production, and DeBraekeleer et al. [13] applied NIR for monitoring the polymorphic conversion.

The on-line study of the state of the process increases the production efficiency and degree of safety. New process automation systems and data visualization tools support this trend. However, the combination of process analytics and data visualization tools creates new tools for understanding the physicochemical phenomena during processing.

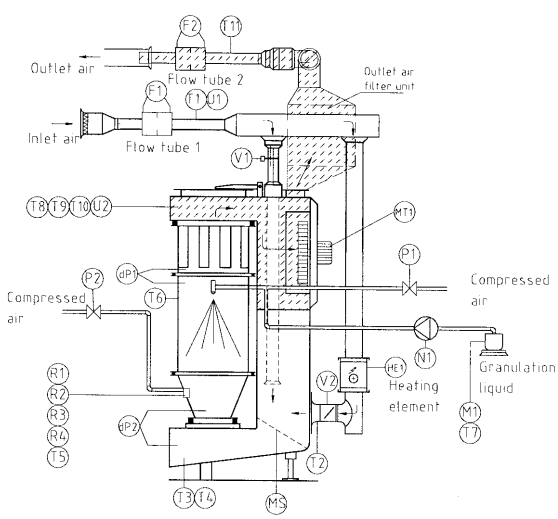

Fig. 1. Instrumentation of fluidized bed granulator: flow rate of air (F), temperature (T), relative humidity (U), NIR reflectance signal (R), pressure difference (dP), pressure (P), mass (M), rotating speed (N), heating element (HE), valve (V), mixing sieve for hot and cold air streams (MS), and motor (M).

Artificial neural networks (ANN) are tools for modeling and visualization of high-dimensional and non-linear data. Back-propagation is a widely used ANN algorithm, e.g. for pharmaceutical product development [14,15] and biopharmaceutical data analysis [16,17]. The self-organizing map (SOM) is an unsupervised ANN method for observing and visualizing high-dimensional data. An overview of process applications has been presented by Simula and Kangas [18]. A general view to process control has also been provided by Tryba [19] and Padgett et al. [20]. The visualization properties have been described by Alhoniemi et al. [21]. Section 2 includes an illustrative description of the SOM algorithm. Further, the SOM was applied to visualize the complex data collected from the FBG process.

2. Experimental

2.1. Granulations

All granulations were performed in a bench-scale fluidized bed granulator (Glatt WSG 5, Glatt GmbH, Binzen, Germany). A schematic illustration of the instrumentation is given in Fig. 1. Measurements included the central properties of process air and product [22]. A NIR set-up was applied for on-line measurement of moisture [6,22]. The supervisory control and data acquisition (SCADA) system was based on the FactorySuite 2000 (v. 7.0, Wonderware Corporation, Irvine, CA). The test material used in granulations was theophylline (BASF, Ludwigshafen, Germany). A batch size of 4000 g was applied. Polyvinylpyrrolidone (PVP) (Kollidon K25, BASF, Ludwigshafen, Germany) was used as a binder. Solutions on purified water were prepared using 20% w/w of PVP. Seven granulations were performed with identical process set values [inlet air temperature 45°C (mixing/spraying phase) and 70°C (drying phase), granulation liquid flow rate 0.1 kg/min, spraying pressure 0.1 MPa]. The flow rate of inlet air was varied between 0.040 and 0.055 m^3/s in order to achieve optimum bed per-

Table 1
Elements of SOM input vector and information included in elements (symbols referring to Fig. 1)

Measurement	Symbol	Information
Process phase indicator	MIX	Discrete variable set by process operator (values 0/1)
Process phase indicator	SPRAY	Discrete variable set by process operator (values 0/1)
Process phase indicator	DRY	Discrete variable set by process operator (values 0/1)
Flow rate of inlet air	F_1	Describes the density of the bed and the heating capacity of inlet process air
Inlet air temperature	T_1	Together with U_1 describes the absolute water content of inlet air
Air temperature after heater	T_2	Indicates possible problems with heater
Air temperature before granulator	T_3	The heating capacity of inlet air can be calculated from T_3 and inlet air water content
Granulation chamber temperature	T_6	Good approximation of granule temperature
Outlet air temperature	T_8	Together with U_2 describes the absolute water content of outlet air
Inlet air relative humidity	U_1	Together with T_1 describes the absolute water content of inlet air
Outlet air relative humidity	U_2	Together with T_8 describes the absolute water content of outlet air
Reflectance at 1813 nm	R_1	Reflectance signal from process at 1813 nm (near-infrared region)
Reflectance at 1998 nm	R_2	Reflectance signal from process at 1998 nm (near-infrared region)
Reflectance at 2136 nm	R_3	Reflectance signal from process at 2136 nm (near-infrared region)
Reflectance at 2214 nm	R_4	Reflectance signal from process at 2214 nm (near-infrared region)
Pressure difference	dP_1	Pressure difference over filter bags, high values indicate blocking of filters
Pressure difference	dP_2	Pressure difference over granules; low values indicate loss of granules or dead zones
Granulation liquid consumption	M_1	Scale information
Apparent water absorbance	AWA_1	Baseline corrected and normalized water absorbance (calculated from R_1, R_2, and R_4)
Temperature difference	T_{diff}	Temperature difference between inlet air and granules, indicating state of granule wetting, and drying

formance. The input vector used for SOM consisted of 20 elements (Table 1).

2.2. Self-organizing maps

The self-organizing map (SOM) is an unsupervised neural network [1]. The SOM can be considered an elastic net with the map of neurons bound to their neighbors by elastic bands. After training, this net travels through densely populated areas of the training data space. This net in an n-dimensional space can be brought back to two dimensions and studied using various visualization tools. This gives information about how the data reside in the n-dimensional space, which is the main benefit of the SOM.

An example of the use of SOM with a simple data set is presented (Fig. 2a). These data contain

two gaussian clusters in three-dimensional space (three variables: Var1, Var2 and Var3). The trained SOM approximates the density distribution with the two-dimensional net (Fig. 2b). The data are visualized in Fig. 3.

The U-matrix is a key to understanding the data (Fig. 3a). The U-matrix presents the elastic SOM grid laid down into a plane. The most important information in the U-matrix is the distances between the nodes of the net. For example, in Fig. 3a, the red color represents the area where the elastic SOM net is stretched between the two data clusters.

The other plots in Fig. 3 show how the SOM occupies the data space. Let us first look at the top cluster. It has high values for variables Var1, Var2, and Var3. For example, the top-right corner of the map has original data points with coordinates of approximately (8, 7, 3). The bottom cluster in general has low values for all three variables. However, the variable Var3 has the relatively high value of zero in most parts of the bottom cluster.

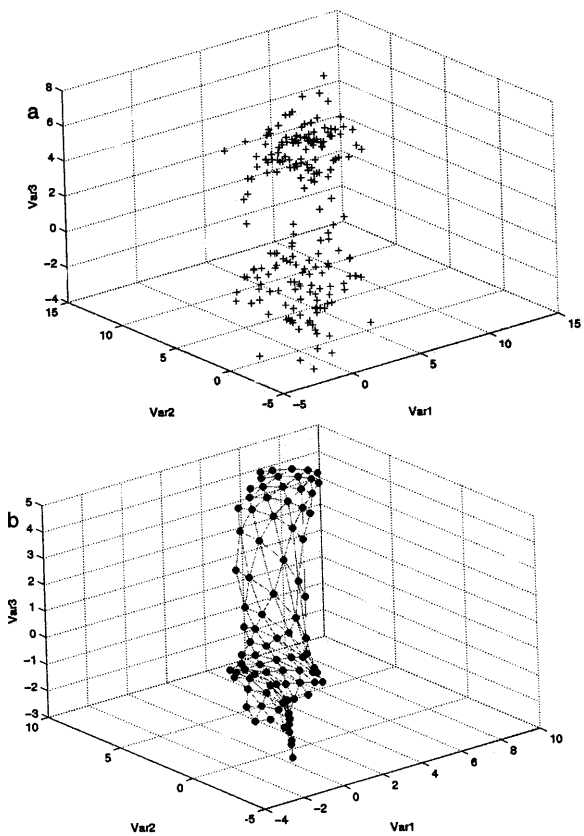

Fig. 2. (a) Demonstration data set in the original data space. (b) Trained SOM with a two-dimensional net.

Similarities and differences exist between the SOM and other data analysis tools. The SOM is similar to the principal component analysis (PCA). Both methods can be used to project a set of n-dimensional data into an m-dimensional space. The PCA is a linear method while the SOM provides a non-linear mapping. The SOM can be related to cluster analysis algorithms. Both SOM and clustering attempt to generate an estimate of the probability distribution of the data. Thus, both methods can be used for classification. The SOM is different from the most usual neural network type, the back-propagation artificial neural network (ANN). The back-propagation ANN is intended to learn the underlying input–output model in data, in a supervised manner. The SOM is intended for the visualization and study of the data and is trained in an unsupervised manner.

Next, the SOM training algorithm is introduced [1]. The text up to the end of this section can be omitted by readers not interested in the technical details. The SOM is trained by introducing the data points one by one to the SOM. In the present study, each point is a vector presenting the values measured at one time. The nodes of the SOM compete for the vector, and the node closest to the input vector is the winner. This winner node and its neighborhood are moved closer to the presented input vector. The nodes of the network gradually learn to represent the training data. Because the neighborhood is taken into account, the properties of adjacent nodes become similar. The map becomes ordered in such a way that clusters with similar properties are located near each other. A part of the SOM with a winner node and a neighborhood around it is illustrated in Fig. 4.

If the input vector is denoted by $x = [x_0\ x_1 \ldots x_{N-1}]^T$ and the location vector of a mapping node by $m_i = [m_{i0}\ m_{i1} \ldots m_{iN-1}]^T$, we can write the algorithm that describes the self-organizing operation:

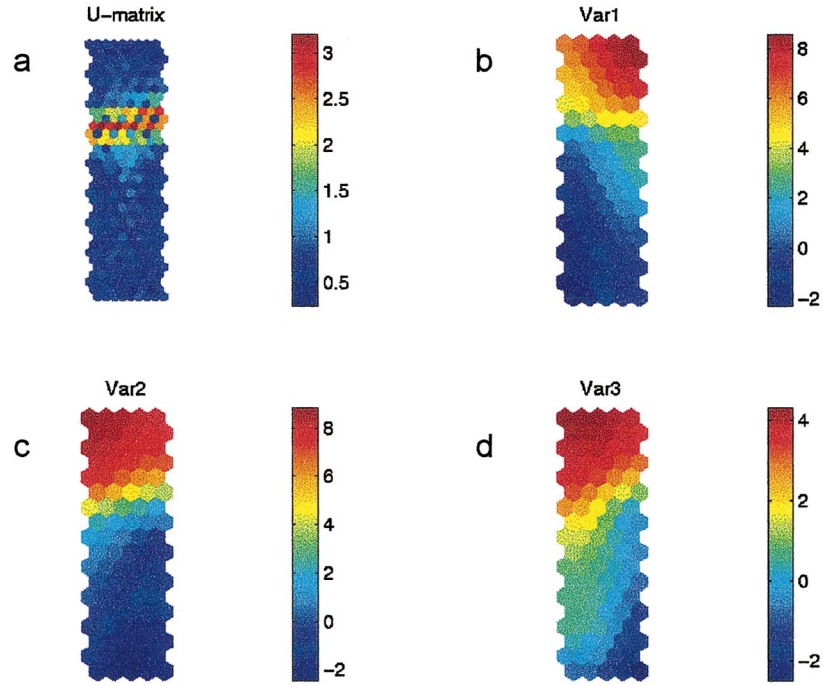

Fig. 3

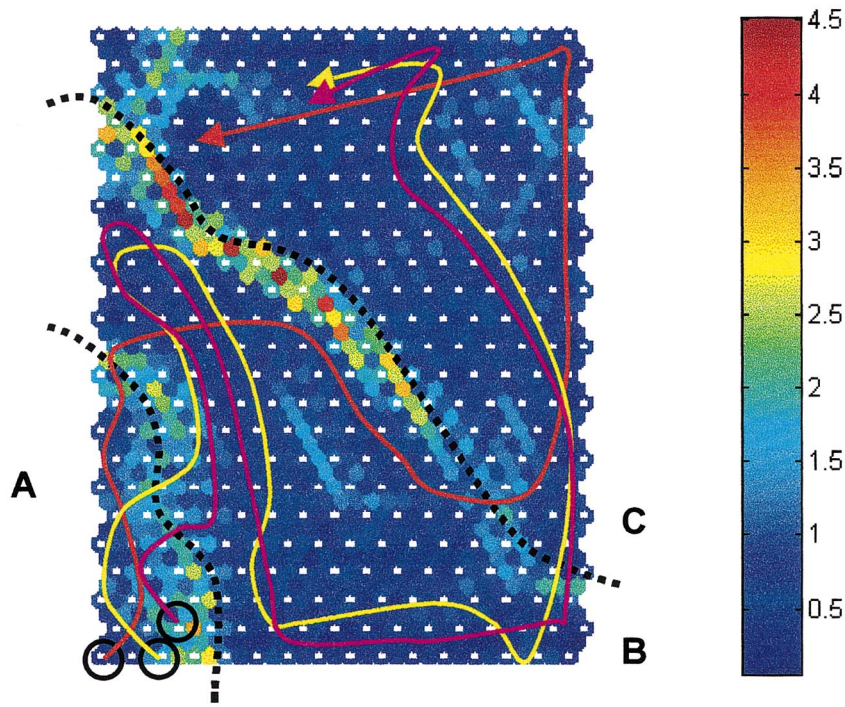

Fig. 6

Fig. 3. The interpretation of the organization of data: (a) U-matrix and (b–d) organization of variable (Var1, Var2, and Var3) information.

Fig. 6. Self-organized map of 20 process measurements from seven granulations; regions A (mixing phase), B (spraying phase), and C (drying phase) indicated by a dashed black line. The color of the hexagon between nodes indicates the distance of the node from its neighboring nodes. Arrows indicate the proceeding of three granulation trace matrices in 2D space (mixing starting from circle

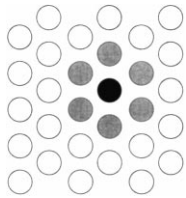

Fig. 4. Example of winner node of SOM with neighbors.

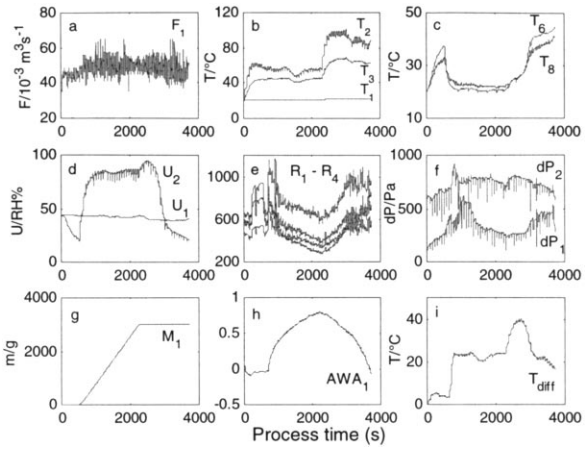

Fig. 5. Traditional trend charts from a typical granulation (symbols referring to Fig. 1): (a) flow rate of inlet air, F_1, (b) inlet air temperatures, T_1–T_3, (c) outlet air temperatures, T_6 and T_8, (d) relative humidities of inlet, U_1, and outlet air, U_2, (e) NIR reflectance signals, R_1–R_4, (f) pressure differences, dP_1 and dP_2, (g) granulation liquid consumption M_1, (h) apparent water absorbance, AWA_1, and (i) temperature difference between granule and inlet air, T_{diff}.

I Initiate the locations of nodes with random values.

II For each vector of the training data compute steps III_a and III_b

III_a Find the SOM node m_c (winner) best matching to the data vector $x(t)$ by searching all nodes m_i:

$$\|x(t) - m_c(t)\| = \min_i \{\|x(t) - m_i(t)\|\} \quad (1)$$

III_b Adjust the locations of the nodes:

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)\|x(t) - m_i(t)\|, \\ \text{for } i \in N_c(= \text{neighborhood}) m_i(t), \\ \text{for all other indices.} \end{cases} \quad (2)$$

In Eqs. (1) and (2), the Euclidean metric can be used as the distance measure.

The parameter $\alpha(t)$ in Eq. (2) is a coefficient that determines how much the winning node and the neighborhood are moved in the direction of the data vector $x(t)$. The parameter $\alpha(t)$ is equivalent to the learning rate used in the back-propagation training algorithm. This $\alpha(t)$ should decrease slowly with time. Initially, $\alpha(t)$ may be chosen as a unity. In the final stages, the value should be less than 0.01. One method for calculating $\alpha(t)$ is presented in Eq. (3), in which T, the total number of iterations, and α_0, the initial value for $\alpha(t)$, are provided by the user of the algorithm.

$$\alpha(t) = \frac{\alpha_0 T}{T + 100t} \quad (3)$$

The neighborhood, N_c, of the winning node, m_c, contains all nodes that are closer than $r(t)$ of the winning node. To ensure good global ordering, the radius, $r(t)$, should initially be more than half the diameter of the network and should slowly decrease with time. In the final stage, a small radius gives the map a good local spatial resolution. In addition to arranging the map topologically, the use of the neighborhood equalizes the number of input vectors classified in each cell.

The third parameter of the training is the number of iteration steps. To reach a good statistical accuracy, the number of training steps must be at least 500 times the number of nodes in the SOM.

3. Results and discussion

3.1. Variables during granulation

The proceeding of a typical granulation is presented in Fig. 5a–i with trend charts. The temperature of the granules is a widely used indicator of the state of the granule wetting and drying phenomena. The temperature of granules was lower (e.g. during the spraying phase and at the beginning of the drying phase, T_6 in Fig. 5c) due to the evaporation of water from the continuous granulation liquid film around the granule. The subse-

quent steps in the drying phase (falling-rate period of drying) were seen as an increased granule temperature. The temperature difference (T_{diff}) between the inlet air and granule temperature has been traditionally applied for the detection of the end-point of drying (Fig. 5i).

In addition, the relative humidity of outlet air changed during processing in a characteristic way (U_2 in Fig. 5d). The absolute water content of the process air stream can be further calculated with temperature data (T_8).

Apparent water absorbance (AWA_1 in Fig. 5h) is a computational NIR signal, which has to be calibrated against a valid reference technique. It can be used to indicate the state of the wetting and drying of granules. All reflectance signals ($R_1 - R_4$ in Fig. 5e) decreased during processing due to increased water absorption, together with a change in particle size and refractive properties of the granules.

The flow rate of inlet air set value was altered during processing in order to achieve optimum bed performance (Fig. 5a). Interruptions due to the filter bag shaking periods were noticed on the flow rate, pressure difference, and reflectance data (Fig. 5a, e and f). Attrition of granules during drying was seen as an increase in pressure difference over filters (dP_1 in Fig. 5f) at the end of the drying phase.

A skilled operator can easily follow this 20-dimensional measurement vector described in Fig. 5 and Table 1. An enormous amount of information is gained, and an expert is needed to visualize the state of the granulation process.

3.2. SOM

For the training and visualization of the SOM, a public domain Matlab toolbox was used [23]. The SOM training parameters are determined automatically by the toolbox based only on training data. The toolbox chooses the number of neurons by the following heuristic formula: $nn = 5 \times dlen^{0.54321}$, where nn is the number of neurons, and dlen is the number of training data samples. Then, the shape of the map is determined by eigenvalues. The two highest eigenvalues of the training data are calculated. The ratio between these two values is set to be the ratio between the side lengths of the SOM grid. The actual side lengths are found by demanding their product to be as close to the desired number of map units as possible. The use of the first two eigenvalues implies that the initialized map grid lies on the PCA plane determined by the first two principal components. From this plane, the map grid evolves according to the SOM training rules.

The training data were created by merging process data collected from seven granulations to one matrix. The size of the SOM created was 16×23 nodes. The training time with a Pentium PC was approximately 40 s.

The trained SOM is presented in Figs. 6 and 7. The former presents the U-matrix showing how the data clusters in the 20-dimensional space. The three clusters corresponded to the mixing (A), spraying (B), and drying (C) phases. A path taken by three granulations through the SOM is shown by an arrow. Each granulation starts from the circle at the bottom-left corner of the map, proceeds through the map by an individual route, and ends at the arrowhead. The identity of the areas of the SOM in Fig. 6 can be deduced from Fig. 7. Discrete variables MIX, SPRAY, and DRY had a value of 1 at these three specific clusters. High values of these variables (red color) can be used for identifying the specific process phases. Further, Fig. 7 visualizes the values of other variables at a glance. For example, the mixing phase is indicated by the low moisture content of mass and can be found at the bottom-left area of the SOM. Similarly, the spraying phase can be found by looking for low temperatures of granules and increasing moisture contents of granules. The drying phase can be found at the top-right part of the SOM.

The evolution through phases of granulation already described in the previous section can be studied using the SOM. The granulations started at the mixing region at the bottom-left corner (Fig. 6, circles at the low-left area of SOM) with a low temperature of inlet air (T_2 and T_3) and low moisture content of mass (AWA_1). The inlet air was heated during the mixing phase, and the trajectory moved upwards. The process proceeded up to the spraying region and continued down-

wards to the right. The most notable change in the process was the increase in the amount of granulation liquid sprayed (M_1) and, thereby, the decrease in granule temperature (T_6) and the increase in the granule moisture content (seen as an increase in apparent water absorbance AWA_1). Then, the process entered the drying region at the bottom-right corner of the SOM. The temperature of the inlet air (T_2 and T_3) increased due to the change in the inlet air temperature set value. As we proceeded to the top of the SOM, different granule drying zones were observed. First, the granule temperature (T_6) increased stepwise. In addition, the region with high values of T_{diff} was observed as a first step of granule drying. At the same time, the moisture content of granules (seen as a decrease in AWA_1) decreased.

The path through the SOM presented the fingerprint of an individual granulation batch. While the overall S-shaped path through the SOM remained unchanged, the granulations proceeded in slightly different paths (Fig. 6). In these experiments, the differences were due to the variation in the process inlet air. One batch (marked with red trajectory) was performed in winter, when the relative humidity of the inlet air was low. The batch proceeded through the regions of SOM with a low value of variable U_1 (Fig. 7). The other two batches (marked by magenta and yellow trajectories) proceeded by different routes due to the higher values of the relative humidity of the inlet air.

A successful granulation can be used to define the optimal path through the SOM (Fig. 8). Then, the deviations from this path can be used to alarm the operator. With the alarm, the operator can be shown the current process state and the variable causing the deviation. In addition, the undesirable regions of the map can be defined (gray areas in Fig. 8). Causes to past granulation failures can be found by determining whether the process state has entered forbidden SOM areas. For example, in the SOM presented above, such an area is in the top-left corner (Fig. 8; I_a) of the drying zone.

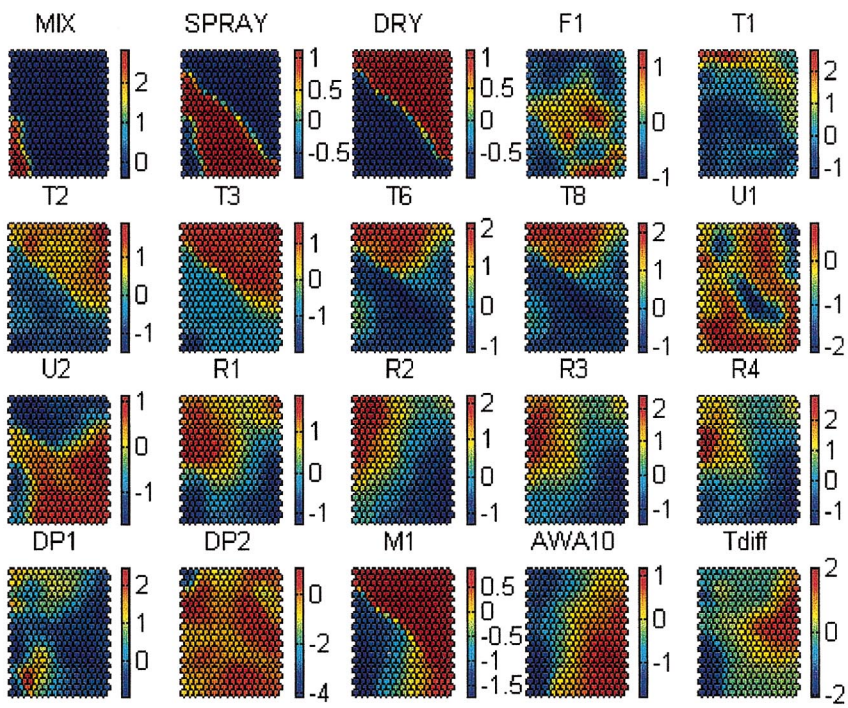

Fig. 7. Interpretation of map information from normalized original variables. The color of the node indicates the level of the individual variable on the specific region of the map.

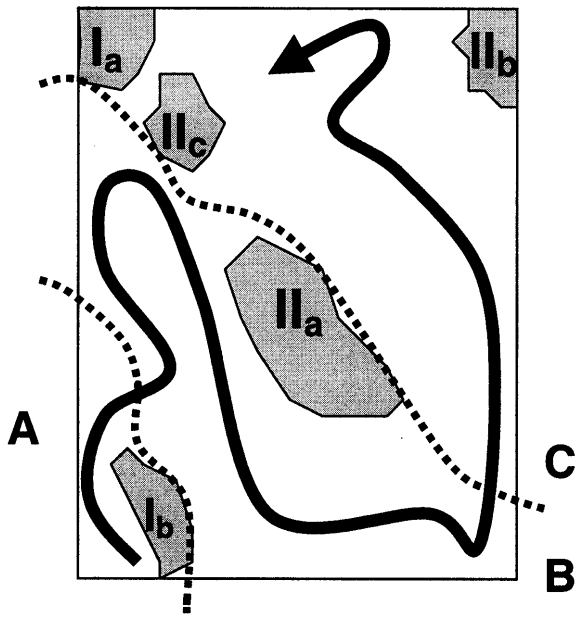

Fig. 8. Proceeding of a successful granulation in self-organized map through regions A (mixing phase), B (spraying phase), and C (drying phase). Gray areas represent the undesirable regions of the map (I_a–I_b, blocking of filters; II_a–II_c, low level of relative humidity of inlet air).

The high value of dP_1 (pressure difference over filter bags) and low value of dP_2 (pressure difference over the air distribution plate and granules) indicate a build-up of material in filters or dead zones (immobile regions) in the granulator. Another area with an indication of filter blinding is in the bottom-left corner (Fig. 8; I_b) of the map. The areas with low values of inlet air relative humidity can also be defined as abnormal regions (Fig. 8; II_a–II_c). This information on optimal and forbidden process states is useful for the process operator. Additional benefit is gained if the operator can be suggested a correcting action. From these results, it is evident that SOM creates a novel tool for visualization of a complex granulation process.

The SOM can be applied to any complex pharmaceutical unit operation with measurements describing the state of the process, e.g. in different stages of solid dosage form manufacture. Development of in-process analytical chemistry, e.g. NIR [2–13], is a powerful tool when combined with data visualization methods. Consider a number of batch processes operated at the same time. The operator can have a SOM presentation of each of the batches and see them at the same time on his/her console. Thus, with one glance, the operator can see if there are any problems in the units. In addition, the operator can quickly determine the reasons for deviations from the optimal path.

Future work should concentrate on conducting similar studies on the mixing, tableting and coating processes, and, further, on the dependences of these stages. The hypothesis is that the process conditions of the predecessor stages have a notable influence on the success of the successor stages. From this study, the whole production chain is optimized.

4. Conclusions

In this study, the self-organizing map was used to visualize the state of a pharmaceutical process. The SOM was able to present clearly the state of the granulation process and the subtle differences between various batches. Several trend charts could be replaced by projecting them into one map. According to these results, it is evident that SOM creates a novel tool for visualization of complex pharmaceutical unit operations.

Acknowledgements

This study was possible thanks to financial support from the Graduate School in Pharmaceutical Research (Ministry of Education, Finland). Lasse Kervinen, M.Sc. (Phys.) (Orion Pharma, Finland) is greatly acknowledged for inspiring discussions.

References

[1] T. Kohonen, Self-organizing Maps. In: Series in Information Sciences, vol. Vol. 30, second ed., Springer, Heidelberg, 1997.

[2] S. Watano, K. Terashita, K. Miyanami, Bull. Univ. Osaka. Pref. Series A 39 (1990) 187–197.
[3] J.G. White, Pharm. Res. 11 (1994) 728–732.
[4] K. List, K.-J. Steffens, Pharm. Ind. 58 (1996) 347–353.
[5] P. Frake, D. Greenhalgh, S.M. Grierson, J.M. Hempenstall, D.R. Rudd, Int. J. Pharm. 151 (1997) 75–80.
[6] J. Rantanen, S. Lehtola, P. Rämet, J.-P. Mannermaa, J. Yliruusi, Powd. Technol. 99 (1998) 163–170.
[7] P.A. Hailey, P. Doherty, P. Tapsell, T. Oliver, P.K. Aldridge, J. Pharm. Biomed. Anal. 14 (1996) 551–559.
[8] S.S. Sekulic, J. Wakeman, P. Doherty, P.A. Hailey, J. Pharm. Biomed. Anal. 17 (1998) 1285–1309.
[9] J.D. Kirsch, J.K. Drennen, J. Pharm. Biomed. Anal. 13 (1995) 1273–1281.
[10] J.D. Kirsch, J.K. Drennen, Pharm. Res. 13 (1996) 234–237.
[11] M. Andersson, M. Josefson, F.W. Langkilde, K.-G. Wahlund, J. Pharm. Biomed. Anal. 20 (1999) 27–37.
[12] S.M. Han, P.G. Faulkner, J. Pharm. Biomed. Anal. 14 (1996) 1681–1689.
[13] K. DeBraekeleer, F. Cuesta Sánchez, P.A. Hailey, D.C.A. Sharp, A.J. Pettman, D.L. Massart, J. Pharm. Biomed. Anal. 17 (1998) 141–152.
[14] A.S. Hussain, X. Yu, R.D. Johnson, Pharm. Res. 8 (1991) 1248–1252.
[15] E. Murtoniemi, J. Yliruusi, P. Kinnunen, P. Merkku, K. Leiviskä, Int. J. Pharm. 108 (1994) 155–164.
[16] P. Veng-Pedersen, N.B. Modi, J. Pharm. Sci. 82 (1993) 918–926.
[17] R.J. Erb, Pharm. Res. 10 (1993) 165–170.
[18] O. Simula, J. Kangas, in: A. Bulsari (Ed.), Neural Networks for Chemical Engineers, Elsevier Science, Amsterdam, 1995, pp. 371–384.
[19] V. Tryba, K. Goser, in: K. Kohonen, O. Mäkisara, J. Simula, N. Kangas (Eds.), Artificial Neural Networks, Proc. ICANN, North-Holland, Amsterdam, 1991, pp. 847–852.
[20] M.L. Padgett, E.M. Josephson, C.R. White, D.W. Duffield, Proc. SPIE — Int. Soc. Optical Eng. 2492 (1995) 562–572.
[21] E. Alhoniemi, J. Hollmén, O. Simula, J. Vesanto, Integ. Computer-Aided Eng. 6 (1999) 3–14.
[22] J. Rantanen, M. Känsäkoski, J. Suhonen, J. Tenhunen, S. Lehtonen, T. Rajalahti, J.-P. Mannermaa, J. Yliruusi, AAPS PharmSciTech 1(2), article 10 (2000), http://www.pharmscitech.com/.
[23] E. Alhoniemi, J. Himberg, K. Kiviluoto, J. Parviainen, J. Vesanto, SOM Toolbox for Matlab, Helsinki University of Technology, Helsinki, 1997. http://www.cis.hut.fi/projects/somtoolbox/.