Journal of Membrane Science 365 (2010) 170–179

Contents lists available at ScienceDirect

Journal of Membrane Science

journal homepage: www.elsevier.com/locate/memsci

Analysis of full-scale membrane filtration processes using econophysics and econometrics

Ying Guo[a], Yong Shi[b], Jame E.T. Moncur[c], Yong Taek Lee[d], Kyoung Wan Kim[e], Albert S. Kim[b,d,*]

[a] School of Accountancy, University of Hawaii at Manoa, 2404 Maile Way, Honolulu, HI 96822, USA
[b] Department of Civil & Environmental Engineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA
[c] Department of Economics and Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI 96822, USA
[d] Department of Chemical Engineering, College of Engineering, Kyung Hee University, Gyeonggi-do 446-701, Republic of Korea
[e] Woongjin Chemical Co., Ltd., 287, Gongdan-dong, Gumi, Gyeongsangbuk-do, Republic of Korea

ARTICLE INFO

Article history:
Received 21 June 2010
Received in revised form 1 September 2010
Accepted 3 September 2010
Available online 16 September 2010

Keywords:
Econophysics
Econometrics
Time series
Membrane filtration
Flux decline
Probability distribution

ABSTRACT

In this paper, econophysics and econometric approaches were employed to investigate statistical and dynamic properties of membrane performance, respectively, using filtration data provided by Woongjin Chemical Co., Ltd., Korea. For the econophysics approach, the filtration index tree was built to comprehensively visualize the closeness between filtration variables, analogous to the hierarchical index tree for the dynamics of stock prices. The logarithmic changes of the filtration variables follow the normal distribution, which implies that the filtration variables themselves are governed by the log-normal distribution due to periodic cleaning and/or seasonal variations. In the econometrics approach, the autoregressive model verified the periodicity of the pressure profile due to cleaning events and natural environmental changes. As developed in this study, the semi-loglinear and autocorrelation models captured pressure growth rates of 1.4% and 0.3 kg/cm^2 per week, respectively. Combined and complementary roles of econophysics and econometrics provided new methodological insights to statistically analyze and further forecast membrane filtration performance in terms of the pressure growth rate in the constant flux operation.

1. Introduction

In the membrane filtration industry, performance forecasting is an important modeling issue, often subject to an empirical analysis of periodically recorded data. Although this approach allows prompt decisions to alleviate flux decline and maintain product quantity for short-term periods, great importance is also given to forecast long-term trends of filtration performance in order to schedule maintenance [1]. Estimation of membrane life is a crucial factor to design, build, and maintain filtration facilities. Depending on the chemical and biological composition of feed water, operational conditions are continually adjusted to achieve production demand for water with desired quality; and autonomous operation can stabilize filtration performance by minimizing chemical and mechanical stresses on membranes. Sophisticated inter-dependencies between filtration variables during operation however preclude capturing the gradual change of filtration performance.

In recent years, several artificial neural network (ANN) methods were developed and applied for phenomenological analyses of measured filtration data without requiring specific transport mechanisms [2–5]. For instance, Hamed et al. [4] studied the dependency between biochemical oxygen demand and suspended solid concentrations through various stages of the treatment process and developed two ANN-based models for predicting the effluent concentrations. Chen and Kim [5] investigated the capability of a radial basis function neural network to predict permeate flux decline in crossflow membrane filtration. However, methodologies developed in this research area are currently limited to short-term predictions of small or lab scale experiments.

To study long-term (monthly and seasonal) changes of membrane performance, this study carefully selected and applied methodologies developed in econophysics and econometrics, analyzed periodically recorded data, and forecasted the performance rate of pressure-driven membrane filtration processes. We used 3.5 years of operating data from Woongjin Chemical Co., Ltd., South Korea, including product water flux, operating pressure, feed concentration, and temperature. A brief review of econophysics and econometrics and their methodologies used in this study follows.

* Corresponding author at: Department of Civil & Environmental Engineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA. Tel.: +1 808 956 3718; fax: +1 808 956 5014.
E-mail address: albertsk@hawaii.edu (A.S. Kim).
URL: http://pam.eng.hawaii.edu (A.S. Kim).

0376-7388/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.memsci.2010.09.009

2. Background

More than a century ago, Italian sociologist and economist Vilfredo Pareto investigated the statistical character of the wealth of individuals in a stable economy and provided the (cumulative) distribution, $y \sim x^{-\nu}$, where y is the number of people having income more than x and ν is an exponent estimated to be 1.5 [6]. Interestingly, Pareto-type power-law distributions were later used in two fields of natural sciences: probability theory [7,8] and phase transitions [9]. In addition to that, stochastic fluctuations – typically described as the random walk stemming from the Brownian motion [10] – are widely used in various disciplines. Bachelier, a French mathematician, was the first researcher who mathematically formalized the random walk, dealing with the pricing of options in speculative markets [11]. Einstein's famous theoretical work on Brownian motion [12] is regarded as the first theoretical description of random walk only in the natural sciences. Later, Wiener established a more rigorous mathematical description of the random walk that is now often called the Wiener process [13,14]. In option-pricing theory, Black and Scholes developed the milestone model using the theory of the random walk, of which step size varies according to a Gaussian probability distribution [15]. Therefore, two of the most widely used mathematical approaches in the natural and engineering sciences, i.e., probability distribution and random walk theories, have close relationships to economics and mathematics. Combining social and natural sciences using mathematics, Stanley and co-workers established econophysics [16–20]. In his pioneering work, Stanley characterized phenomenological dependences of growth on company size using data from 1975 to 1991 for all publicly traded US manufacturing companies [16]. In general, econophysics employs mathematical tools developed in theoretical statistical physics. Using fundamentals of statistical mechanics, researchers investigated statistical properties of sophisticated changes in economic systems and described the phenomenological rules in financial phenomena such as the stochastic process of financial asset price changes [21–24], income distribution of firms [16,25], and the time correlation of financial fluctuations [26]. To solve econometric problems, the maximum entropy method [28,27] and Bayes' rule [29,30] were recently combined into the method of maximum relative entropy [31,32], which has advantages of assigning as well as updating probabilities when new information is provided in the form of posterior constraints. Bayesian perspective on how the probability distribution links between statistical physics and other sciences can be found elsewhere [31,32].

Econometrics has a longer history than econophysics. The term "econometrics" was credited to Ragnar Frisch, who developed and applied dynamic models to the analysis of economic processes [33] and received the first Nobel Prize in Economic Sciences in 1969 with co-recipient Jan Tinbergen. After Frisch and Tinbergen, the Nobel Prizes in Economic Sciences was awarded four more times to econometrics researchers. Econometrics is a statement of economic theory in mathematical form used to develop statistically testable hypotheses including methods for identification and estimation of simultaneous economic variables. The basic statistical method used by econometricians includes regression analysis using data sets with inter-dependent variables [30]. Although parameter estimation of regression models using cross-sectional data (i.e., data collected by observing many variables at the same time) is often viewed as a primary goal of econometrics, of equal importance is the production of good methods of economic forecasting, which can be classified into two broad categories as follows: first, casual forecasting models that can forecast a dependent variable if the associated values of the independent variables are given after parameter estimation of models; and second, time series models that can provide some guidelines as to what to expect in the future using data history. Generally, econometrics is a unification of statistics, economic theory, and mathematics to understand the quantitative relations in modern economic life. Therefore, these methods can be widely used in various disciplines including engineering sciences.

3. Econophysics approach

In this section, we explain how to develop correlations and taxonomy of stock-price of companies, and apply the same method to membrane filtration processes by replacing the stock prices with filtration data.

3.1. Correlations and taxonomy of stock prices

Microscopic transport equations, which describe interactions among various economic entities, are barely known or only partially understood. Methodologies and concepts developed in statistical physics were employed to understand complex financial phenomena [34] such as stochastic dynamics, random fluctuations, short- and long-range correlations, self-similarity, length/time scaling, and renormalization. To understand simultaneous dynamics of stock pairs, econophysics measures similarities and dissimilarities along with synchronous variations of stock pairs without postulating specific mechanisms. A simple method to quantify correlations between two stocks evolving in a synchronous fashion is as follows.

If a time series of the daily closure price of stock i, i.e., $P_i(t)$, is available, its logarithmic change $S_i(t)$ can be calculated as

$$S_i(t) = \ln \frac{P_i(t+\Delta t)}{P_i(t)} \approx \frac{\Delta P_i(t)}{P_i(t)} = R(t) \tag{1}$$

where the time interval Δt is a day, $\Delta P_i(t)$ is the stock price difference between two consecutive days, and $R(t)$ is defined as return or rate of growth. It should be noted that the logarithmic functional form of Eq. (1) stems from the fundamental concept of statistical mechanics, entropy [35,36], more specifically relative entropy [28,27,37] which measures the (abstract) distance between two probability distributions, $P_i(t)$ and $P_i(t+\Delta t)$. According to logarithmic properties, $S_i(t)$ reflects the daily percent-change (rate) of stock price i at time t, and it is equivalent to the return $R(t)$ if the daily variation $\Delta P_i(t)$ is small. Now, the correlation coefficient ρ_{ij} between S_i and S_j is derived as

$$\rho_{ij} = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\langle S_i^2 - \langle S_i \rangle^2 \rangle \langle S_j^2 - \langle S_j \rangle^2 \rangle}} \tag{2}$$

where the angular brackets indicate the average over a period of time. The range of the correlation coefficient varies from −1 (completely anti-correlated) to +1 (completely correlated), and $\rho_{ij} = 0$ indicates the independence between two stock dynamics [38].

By building the correlation coefficient matrix based on graph theory [39], one can measure the similarity between two stock portfolios by calculating their relative distances and form groups of closely correlated entities [40]. Let us define a new dimensionless variable

$$\tilde{S}_i = \frac{S_i - \langle S_i \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}} \tag{3}$$

where $\tilde{S}_i$ is a normalized S_i, i.e., the difference between the logarithmic change of a daily stock price i and its mean, divided by its standard deviation. The Euclidean distance d_{ij} between $\tilde{S}_i$ and $\tilde{S}_j$ can be obtained from the Pythagorean relation:

$$d_{ij}^2 = ||\tilde{S}_i - \tilde{S}_j||^2 \tag{4}$$

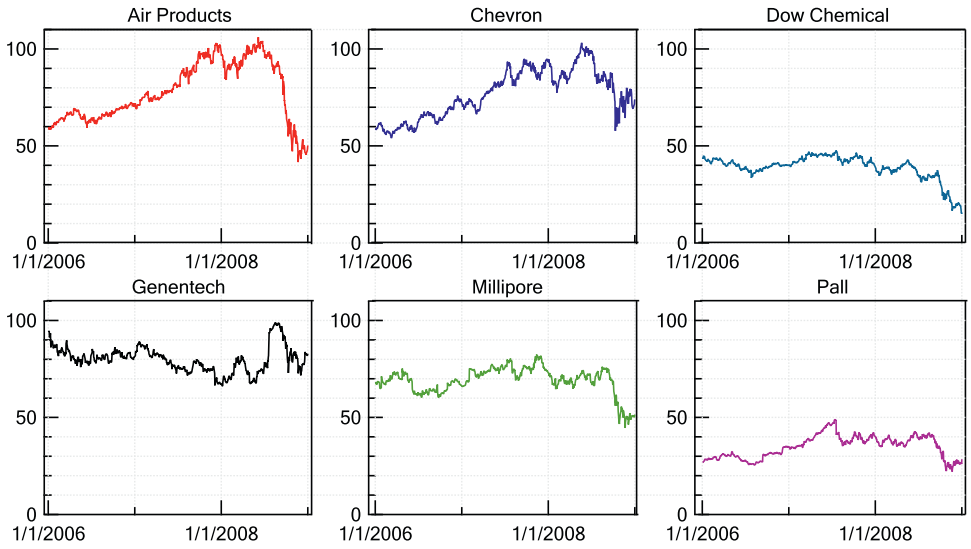

Fig. 1. Daily closure stock prices (US$) of the six selected sponsors, from 2006 to 2008, for 2009 NAMS conference.

Then, one can alternatively define the correlation coefficient as

$$\rho_{ij} = \langle \tilde{S}_i \tilde{S}_j \rangle = \frac{1}{n} \sum_{k=1}^{n} \tilde{S}_{ik} \tilde{S}_{jk} \tag{5}$$

and represent the Euclidean distance as

$$d_{ij} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (\tilde{S}_{ik} - \tilde{S}_{jk})^2} = \sqrt{2 - 2\langle \tilde{S}_i \tilde{S}_j \rangle} = \sqrt{2(1 - \rho_{ij})} \tag{6}$$

where $\tilde{S}_{ik}$ represents $\tilde{S}_i$ (of company i) at time (day) k in a period of n days, having a property of $\langle \tilde{S}_{ik}^2 \rangle = \langle \tilde{S}_{jk}^2 \rangle = 1$. The Euclidian distance in Eq. (6) has the following properties [38]: (i) $d_{ij} = 0$ if $i = j$, (ii) $d_{ij} = d_{ji}$, and (iii) $d_{ij} \leq d_{ik} + d_{kj}$. If the property (iii) can be replaced by a stricter criterion, $d_{ij} \leq \max\{d_{ik}, d_{kj}\}$, the Euclidean distance obeys the ultrametric inequality in a ultrametric space [41]. The ultrametric space provides a convenient reference frame to understand a hierarchically structured complex system endowed with ultrametric distances [42]. Among all the possible ultrametric spaces, we chose the subdominant ultrametric space, which is associated with the Euclidean distance metric d_{ij} using Kruskals algorithm [39,43].

Now we illustrate how to apply the above-mentioned econophysics approach to build the correlation matrix and hierarchical index tree (HIT) of stock prices among six publicly traded companies. Although an unlimited number of arbitrarily selected companies in various industries can be considered for the econophysics analysis [16], we selected the following six companies who were financial sponsors of the North American Membrane Society (NAMS) 2009 Annual conference (held in Charleston, SC, USA) where the current work was partially presented: Air Products (Detroit, USA), Chevron (San Ramon, USA), Pall (Port Washington, USA), Genentech (South San Francisco, USA), Dow chemical (Midland, USA), and Millipore (Billerica, USA). We chose these companies because our primary goal is to analyze and forecast membrane filtration performance, and sales products of the six engineering companies are related to membrane filtration processes to a certain degree. Variations of their stock prices were influenced by several major factors, of which in-depth investigations are truly in fields of industrial economics and business management. Other NAMS 2009 sponsors such as foreign companies and US federal agencies were excluded in this study because of unavailability of comparable stock data to those of the six above-mentioned companies.

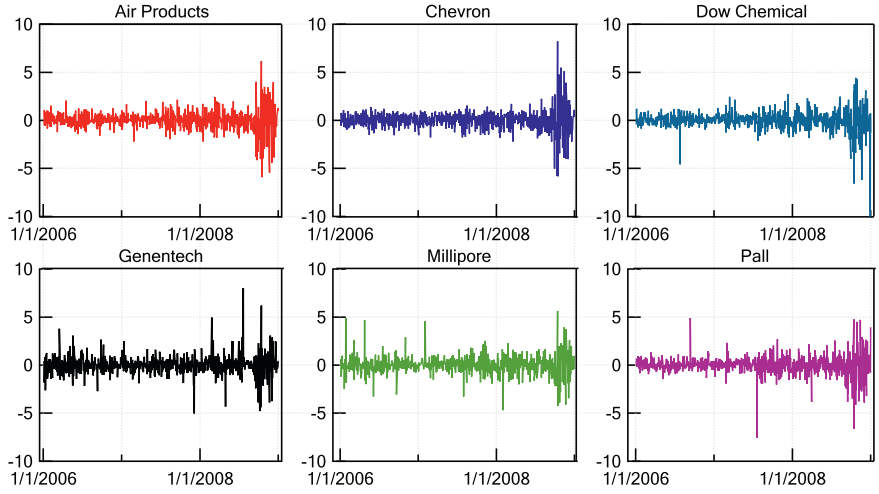

Fig. 2. Logarithmic changes of daily closure stock prices (US$) of the six selected sponsors, from 2006 to 2008, for 2009 NAMS conference.

Table 1
Density correlation matrix (ρ_{ij}) of the six selected NAMS sponsors: DC (Dow Chemical), AP (Air Product), CV (Chevron), GT (Genetech), MP (Millipore), and PL (Pall).

ρ_{ij}	DC	AP	CV	GT	MP	PL
DC	1.000	0.646	0.656	0.305	0.508	0.600
AP		1.000	0.715	0.360	0.600	0.689
CV			1.000	0.362	0.539	0.637
GT				1.000	0.374	0.330
MP					1.000	0.579
PL						1.000

Their stock prices data (US$) were obtained from the database of Mergent Online (http://www.mergent.com/, Mergent Inc., New York, USA), a leading provider of business and financial data on global publicly listed companies. We did not include foreign sponsors, because their stock information was not available from the same data source.

The three-year (from 2006 to 2008) daily close stock prices for these companies are shown in Fig. 1. Consistent increase was observed in stock prices of Air Products, Chevron, and Pall in 2006 and 2007 with (rapid) declines in the middle of 2008. Dow Chemical and Millipore have relatively plateau-shape profiles, gradually decreasing in 2006–2007 before drops in 2008. Genentech shows a unique trend that gradually decreases from 2006 to 2007, followed by a sharp peak in mid-2008. Fig. 2 shows similarities of logarithmic stock-price changes for two groups: Air Products/Chevron/Pall and Dow Chemical/Millipore. As expected, Genentech's logarithmic profile is dissimilar to those of the other companies. The profiles shown in Fig. 2 resemble 1D random walks [44], and statistical analysis should provide better understanding than visual comparisons.

These dynamic trends can be quantified with correlation coefficients used to calculate the Euclidean distances. By running statistical software E-views (Quantitative Micro Software, LLC, CA, USA), we obtained the correlation coefficient matrix ρ_{ij} for the six companies listed above, and the results are shown in Table 1. It should be noted that all the correlation coefficients are positive numbers. A negative correlation coefficient implies anti-correlation, i.e., the opposite trend that one increases and its pair decreases. The synchronous patterns of all six companies have a certain degree of similarity. Using the data in Table 1, we computed the Euclidean distances d_{ij} as shown in Table 2. For positive correlation coefficients, standard values of the Euclidean distance are as follows: $d = 1.414$ for $\rho = 0.0$, $d = 1$ for $\rho = 0.5$, and $d = 0$ for $\rho = 1.0$. Specifically, a Euclidean distance greater than 1.0 indicates a correlation coefficient below 0.5, which is the case for Genetech when correlated with the other five companies.

From the Euclidean distance matrix in Table 2, we obtained the HIT, shown in Fig. 3, using the following method. First, we found a stock pair of the shortest distance, which is 0.754 between Chevron and Air Products and connected them as the closest group. The second-shortest distance is between Air Products and Pall at 0.789, which is quite close to that of Chevron and Air Products. So, we connected Pall to the group that Air Products is associated with. We continued to find the shortest distances until all of the six companies were grouped. As the HIT in Fig. 3 indicates that the strongest correlation is between Chevron and Air Products, their daily closure stock price curves shown in Fig. 1 closely resemble each other. Overall, Chevron, Air Products, Pall, and Dow Chemical are closely related as an interacting group within Euclidean distances of 0.790 ± 0.038. All four companies are involved in the chemical engineering industry. Similar material supplies and market demands probably contribute to the similarity of their stock price dynamics. However, Genentech focuses on the biotech business so this industrial dissimilarity could contribute to Genentech's larger distances from other companies.

Table 2
Euclidean distances (d_{ij}) of the six selected NAMS sponsors, calculated using the correlation density values of Table 1: DC (Dow Chemical), AP (Air Product), CV (Chevron), GT (Genetech), MP (Millipore), and PL (Pall).

d_{ij}	DC	AP	CV	GT	MP	PL
DC	0.000	0.842	0.829	1.179	0.992	0.894
AP		0.000	0.754	1.132	0.895	0.789
CV			0.000	1.130	0.960	0.852
GT				0.000	1.119	1.157
MP					0.000	0.918
PL						0.000

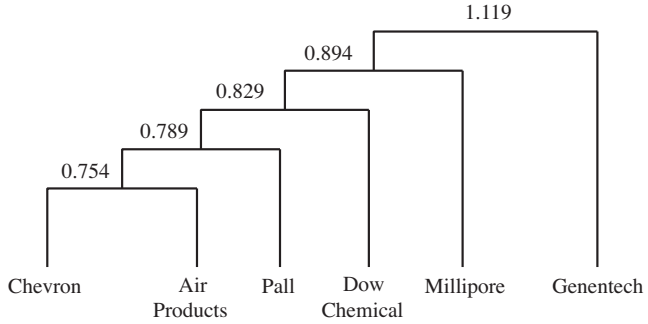

Fig. 3. The hierarchical index tree (HIT) of the six selected NAMS sponsors.

In this section, we defined the dimensionless logarithmic change of a stock price and calculated the correlation density and Euclidean distances using standard techniques suggested by econophysics. Results shown here are based on publicly known stock information of the six selected sponsors of 2009 NAMS conference. Primary reasons for the stock price variation of each company over a short time need to be sophisticatedly investigated using social, political, and perhaps cultural reasons in addition to national and global demands of filtration-based water production. However, we focused on how econophysics analyzes dynamic stock trends without deeply looking into micro-scale economic changes because scrutinization of independent governing factors and their economic significances is out of the scope of this study. Now, we take the exactly same strategy to investigate correlations and taxonomy among filtration variables which were recorded daily in an actual membrane facility operated by Woongjin Chemical.

3.2. Filtration variables

3.2.1. Correlations and taxonomy

In this section, we use the above-mentioned method to build a HIT of filtration variables, designated as a filtration index tree (FIT). Data obtained from Woongjin Chemical include permeate flow rate, operating pressure, feed conductivity, and feed temperature recorded from November 16, 2000 to May 22, 2004, at a membrane facility located in the city of Ulsan, South Korea. In order to use the original data directly, we equated feed conductivity and product flow rate with concentration and permeate flux, respectively. These direct conversions are permissible because the actual feed concentration monotonically increases with conductivity, and the total surface area of a number of membrane modules is fixed during operations. More important, dimensionless quantities are used in econophysics approaches, and the fundamental transport mechanisms do not depend on variable units.

Fig. 4 shows synchronous variation of the four operational quantities. Concentration profile indicates semi-periodical variations, and temperature seems to follow a seasonal concentration change. The temperature control prevented the feed temperature from decreasing to less than 15 °C. The initial plateau of the temperature profile may be ascribed to ill-functioning temperature sensors

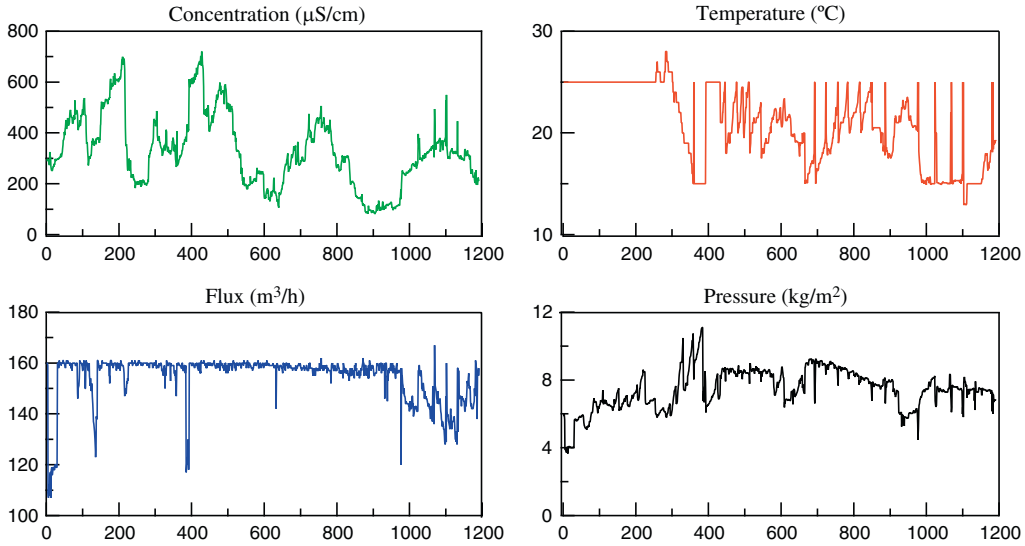

Fig. 4. Three-and-half-year membrane filtration operational data for permeate flux, pressure, feed concentrations, and temperature from Woongjin Chemical Co., Ltd. (Seoul, Korea).

Table 3
Euclidean distances (d_{ij}) of the operational variables from Woongjin Chemical Co, Ltd. (Seoul, Korea): concentration, flux, pressure, and temperature.

d_{ij}	Concentration	Flux	Pressure	Temperature
Concentration	0.00	1.118	1.623	0.848
Flux		0.00	1.275	1.126
Pressure			0.00	1.738
Temperature				0.00

or improper management of data. The permeate flux was stable except a few extreme events, especially during the initial filtration period. Long-term variation of applied pressure is observed, but the trend does not seem to follow the other variables.

For econophysics analysis, we treated each filtration variable as a stock price, used the same methods discussed in Section 3.1, and calculated the Euclidean distances shown in Table 3. The FIT for the four filtration variables is built and shown in Fig. 5. The temperature and the feed concentration have the closest distance of 0.848 as indicated in Table 3 and therefore form the first (closest) group in Fig. 5. The second closest distance, 1.118, is between the concentration and flux, which is very close to that between the temperature and flux, 1.126. Using either distance gives the same tree structure among the temperature, concentration, and flux because, in both cases, the flux is connected to the primary temperature–concentration group. Finally, the flux has a distance of 1.275 to the pressure.

The physical meanings are as follows. The closest distance of concentration to temperature stems from the seasonal variation of feed water quality. The four seasons in South Korea may contribute to the similarity between transient variations of the concentration and temperature of feed water. Darcy's law can contribute to explaining the permeate flux J as influenced by temperature–concentration group:

$$J = \frac{\Delta P - \Delta\pi(C)}{\mu(T)R_m} \tag{7}$$

where ΔP is the pressure, $\Delta\pi(C)$ the osmotic pressure difference as function of the feed concentration C (measured as feed conductivity), $\mu(T)$ the feed water viscosity varying with temperature T, and R_m the membrane resistance. Because the osmotic pressure is a monotonically increasing function of concentration C, when the feed concentration increases flux decline occurs. In addition, the kinematic viscosity of pure water increased 28% from 25 °C to 15 °C, which indicates that the seasonal variation remarkably influences the long-term variation of the permeate flux. Fig. 5 therefore captures the important physical characteristics to express that the feed concentration and temperature influence the permeate flux through the osmotic pressure and kinematic viscosity, respectively. Now it is worth noting that pressure is least correlated to the other quantities. This is because the filtration system was operated in the constant flux mode. During the period that the feed concentration and temperature were fluctuating on a daily basis, the trans-membrane pressure was autonomously set to keep the permeate flux as a pseudo-constant. The constant flux operation mode made the pressure passively evolve in response to changes of C and T through J. In general, when a number of variables are monitored during large-scale membrane filtration processes, the FIT analysis can group several key variables that influence filtration performance in terms of flux or pressure. This classification may help identify critical components causing serious concentration polarization followed by membrane fouling phenomena.

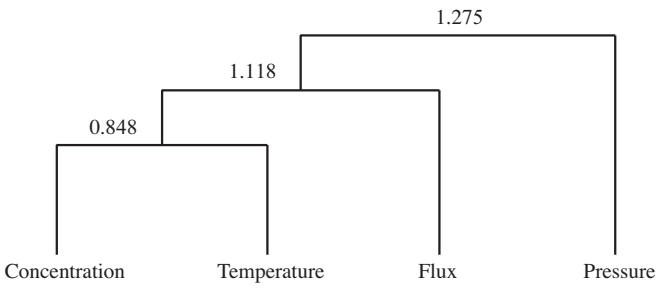

Fig. 5. The FIT of permeate flux, operating pressure, feed concentration, and temperature, obtained from Woongjin Chemical Co. Ltd. (Seoul, Korea).

3.2.2. Probability density function

The FIT in the previous section provided information about dynamic similarities between filtration variables, basically measured as correlation coefficients. As indicated, the logarithmic change of the filtration variables is equivalent to the percent-rate of the individual variable. Now it is interesting to see how frequently a certain percent change happens during the entire filtration period. In the financial market, the probability density function (PDF) has been widely used when stochastic processes of price changes are

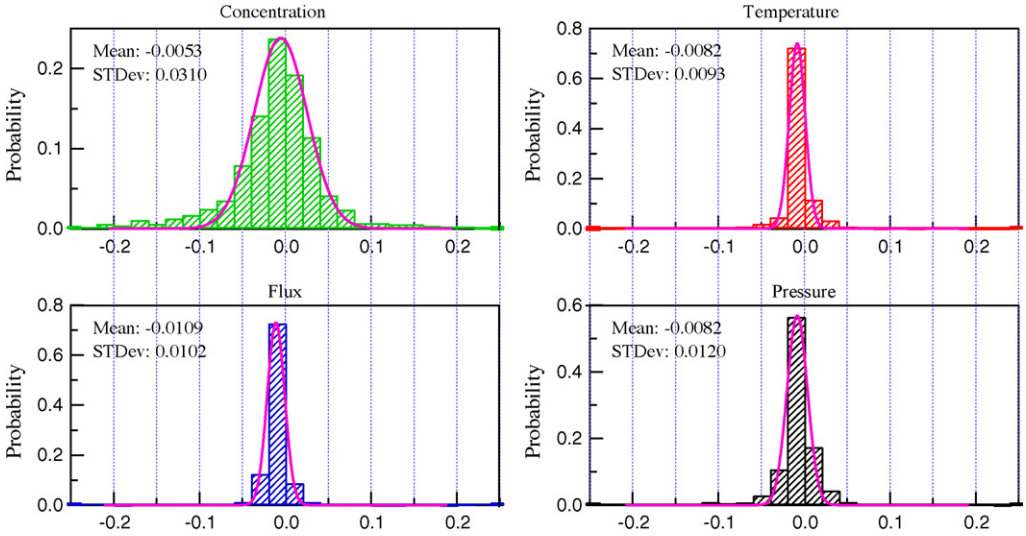

Fig. 6. The logarithmic change-probability density functions (LC-PDFs) for concentration, temperature, flux and pressure.

investigated using econophysics [8,45–47]. In the same line, we built the logarithmic change-probability distribution function (LC-PDF) of the above-mentioned operating variables and investigated the statistical properties of the dynamically changing filtration performance.

Fig. 6 shows LC-PDFs of the four filtration variables. The concentration logarithmic change has the largest standard deviation of 0.031 while others have deviations around 0.01. The wider distribution of the concentration logarithmic change stems from the seasonal and yearly variations with sudden large increases and decreases as shown in Fig. 4. Although the temperature is closely related to the concentration as shown in the FIT of Fig. 5, artificial temperature control maintains the temperature in the range of 15–25 °C and therefore provides the smallest standard deviation of the logarithmic change. It is interesting to compare the LC-PDF of J and ΔP in Fig. 6, which have slightly opposite trends: LC-PDF of J is slightly left-weighted and that of ΔP right-weighted. Because of concentration polarization followed by membrane fouling during normal operation, the permeate flux gradually declined due to the influences of daily fluctuating concentration and temperature. To bring the flux back to a constant pre-set value, the pressure needed to be increased, roughly the same frequency of flux change. After each cleaning event, the permeate flux suddenly rose with the less hydraulic resistance in comparison to that before the cleaning. The constant flux operation led to pressure reduction to mitigate sudden flux increases, primarily to maintain a flow rate of 160 m^3/h as indicated in Fig. 4. Because the cleaning is intermittent, occurring every one to two months, the frequency of the flux decline (i.e., pressure increase) is higher than that of the flux rise (i.e., pressure decrease) over the total operational duration. This explains the LC-PDF shapes of J and ΔP in Fig. 6.

With acceptable tolerance, all PDFs show narrow, symmetric shapes. This feature tells us that a filtration variable has similar probability to increase or decrease with a percentage at given time t. Therefore, the daily percent-change rates of C, T, J, and ΔP can be treated as (quasi-) random walk processes, i.e., Markov chains: today's filtration performance is primarily affected by yesterday's filtration status with negligible influences from the operational history. This random walk is approximately analogous to the molecular random walk in the phase space [48,49] especially when returns of Eq. (1) are small, because the daily change of a filtration variable can be treated as its velocity. The original filtration variables (C, T, J, and ΔP) are guided by mutual influences of underlying microscopic transport mechanisms. In other words, the normal (Gaussian) shape of the four LC-PDFs indicate C, T, J, and ΔP (not their log-changes) follow the log-normal distribution of a right-skewed tail. Intermittent occurrences of large C, T, J, and ΔP imply that their sudden increases and decreases may be due to (natural) seasonal variations of C and T and semi-periodic cleaning events in terms of J and ΔP.

In addition, we ran the augmented Dickey-Fuller (ADF) test on the operational data series. The underlying criteria of an ADF test is that if the series is integrated, then the lagged level of the series $X_i(t-1)$ provides no relevant information in predicting the change in $X_i(t)$ besides the one obtained in the lagged changes $\Delta X_i(t-k)$ for $k = 2, \ldots, p$ where p is the number of augmenting lags [30]. Here, X_i indicates the time series data of the concentration, temperature, flux, and pressure (not the logarithmic changes). Once a value of the ADF test statistic, denoted as τ, was computed, we compared it to the relevant critical value from the standard Dickey-Fuller test. The τ-values of the four filtration variables were calculated as follows: concentration (−2.79), temperature (−5.65), flux (−5.80), and pressure (−5.05). Except for the concentration time series, all the τ-values are less than the critical value, −3.41 at 95% confidence level. This indicates that the time series data are stationary (except concentration), implying that we have sufficient evidence here to draw a conclusion on its stationarity. As indicated, the τ value of concentration is close to but larger than the critical value. However, the daily logarithmic change of concentration is considered stationary on the basis of its τ-value of −41.70, from which we can build a model to connect operational variables. The τ values of the logarithmic change of temperature, flux and pressure are −53.0, −30.4 and −37.6, respectively, and the critical value for the Augmented Dickey-Fuller unit root tests is −2.86 at 95% confidence level. This verifies without losing generality that logarithmic changes of all four filtration variables can be treated as Markov Chains.

Table 4
Coefficients of the autoregression model, AR(200), of Eq. (8).

P_0	α_1	α_7	α_{36}	α_{37}	α_{49}	α_{54}	α_{55}	α_{66}	α_{81}	α_{91}	α_{126}	α_{162}	α_{180}	α_{181}	α_{196}
0.3028	+0.8776	+0.0835	−0.1405	+0.1584	0.1245	0.1263	−0.1889	−0.1148	0.0962	−0.0939	−0.1134	0.0904	0.1290	−0.0975	−0.0920

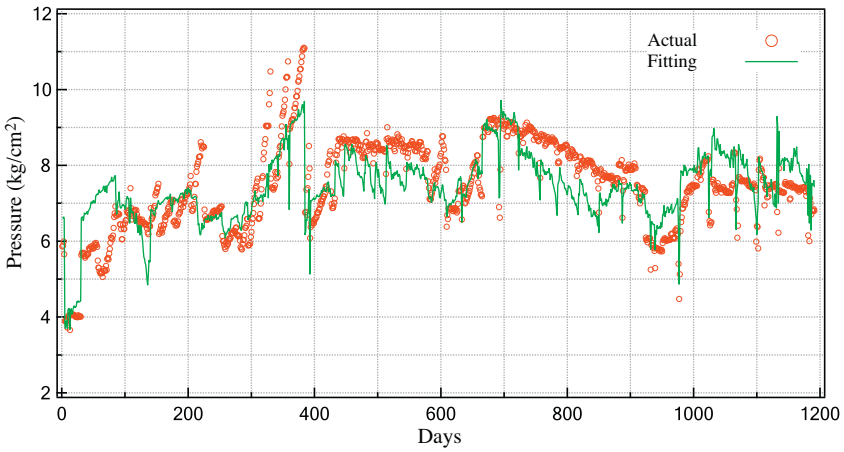

Fig. 7. The comparison of simulated pressure with actual data based on the model described by Eq. (9). The circles denote the actual operational data and the line represents the result from fitting. Parameters used in Eq. 9 are $\beta_P = -5.304$, $\beta_J = 1.470$, $\beta_C = 9.362 \times 10^{-2}$, $\gamma = 3.206 \times 10^{-2}/^{\circ}$C, and $\omega = 2.031 \times 10^{-3}$/day, which are all significant with $R^2 = 0.6190$.

4. Econometric approach

The econophysics approach used in the previous section treated each value set of C, T, J and ΔP (and their logarithmic changes) as independent statistical events at time t. In this section, we use the econometric approach to analyze asymptotic dynamics of filtration variables in the operational history. The autoregressive and autocorrelation models are employed to analyze filtration time series and forecast the future performance in terms of pressure growth rate.

4.1. Periodicity of pressure profile

As filtration variables, including the pressure, follow a (quasi-) lognormal distribution, it would be worth investigating the periodicity of the pressure time series. As discussed in Fig. 5, the pressure passively responds to natural variations of C and T to maintain the permeate flux as a quasi-constant. Therefore, the pressure can be a measure of membrane fouling and an indicator for the next membrane cleaning. To study effects of the pressure of past p days on the current pressure, we used the autoregressive model, AR(p):

$$\Delta P(t) = P_0 + \sum_{i=1}^{p} \alpha_i \Delta P(t-i) + e_t \tag{8}$$

where P_0 is the constant swift term, α_i is the coefficient of time-lag i, and e_t is the white noise. The order p is set to 200 days (arbitrarily chosen) to observe long-term influences, and all 201 coefficients (including P_0 and $\alpha_1 - \alpha_{200}$) are calculated using software E-views. Table 4 shows selected coefficients of important values, and all other α_i's (not shown) are of much less magnitude than those of Table 4. The constant swift P_0 (= 0.3028 kg/cm^2) is negligible in comparison to the average pressure value over the entire three and a half years, as calculated, $\langle \Delta P \rangle$ = 7.460 kg/cm^2. The immediate response α_1 (= 0.8776), close to 1.0, indicates the significant influence of the pressure from the day before on the current pressure. A weekly variation, described as α_7 (= 0.0835), is observed to be much less significant than the immediate influence, α_1, but comparable to coefficients of longer time-lags shown in Table 4. Approximate periodic correlations are indicated by the coefficient indexes for the past one month (36 and 37), followed by two (49, 54, 55, and 66), three (81 and 91), four (126), five (162) and six (180, 181, and 196) months. This emphasizes monthly correlations due to cleaning events and natural seasonal variations in C and T, although indexes of coefficient α_i only roughly match multiples of 30 days. One-month correlation occurs about 5 weeks (i.e., days 36 and 37), and a two-month correlation is spread from days 49 to 66 (56 ± 7 days on the average). This implies that the cleaning was probably irregularly performed due to the natural fluctuation of input conditions. (See the next section for details about membrane cleaning.) Seasonal correlation is clearly shown by coefficient indexes, 81 and 91 for three months, and 181 and 196 for six months, having slightly less significance than those of other monthly correlations. Offsets of one or two weeks were observed in four- and five-month correlations of days 126 and 162, respectively, and interestingly the coefficient of the five-month correlation α_{162} has the least magnitude among all the monthly correlations.

4.2. Models for membrane performance evolution

4.2.1. Semi log-linear model

To mimic basic filtration behavior, we begin with the fundamental transport Eq. of (7), Darcy's law, and suggest the following semi-loglinear regression:

$$\ln(\Delta P) = \beta_P + \beta_J \ln(J) + \beta_C \ln(C) - \gamma T + \omega t + \epsilon \tag{9}$$

where β_P, β_J, β_C, γ and ω are coefficients to be calculated (see Fig. 7), ϵ is the disturbance, and t is the elapsed time. Observations about C, T and J are not assumed to convey any information about the expected value of the disturbance, defined as the difference between the model prediction and real observation. Membrane cleanings were performed intermittently when the permeate flux fell below a desired value and so the pressure rapidly and consistently increased. Specific dates of membrane cleaning events with detailed guidelines were not provided. The time series data of J and ΔP show sudden large increases and decreases, which seem to be repeated about every one or two months. We looked at the ΔP time series data and reset time t to zero when a sudden drop in ΔP is seen, regarding each as a cleaning event. In this way, the average pressure increase between two consecutive cleaning events can be estimated. The gradual increase of this rate over a longer duration may explain membrane aging but requires much advanced statistical analysis. Eq. (9) represents the pressure (a dependent variable) using the independent variables of J, C, T and t. Although Darcy's law uses net pressure as the difference between the applied pressure and the (feed) osmotic pressure, we assumed separate influences of J and C in Eq. (9) because of the low feed conductivity. There are several versions of empirical equations to correlate the conductivity to concentration: in most cases, one μS/cm is roughly equivalent to 0.5 ppm or below. Since the water viscosity expo-

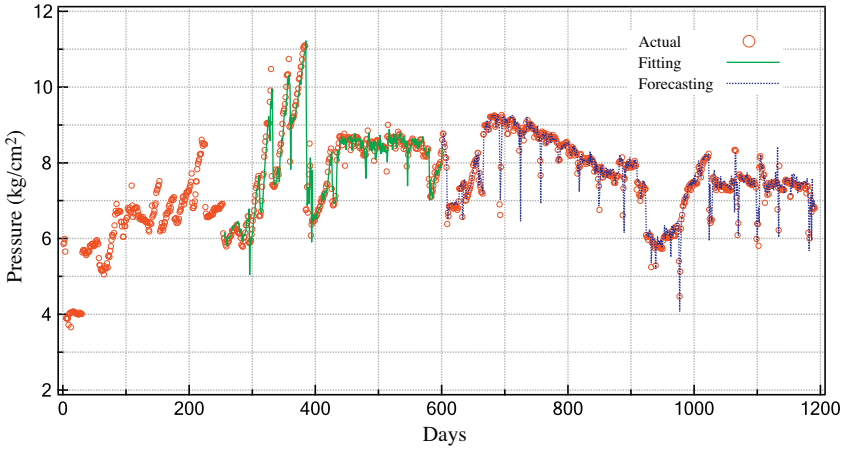

Fig. 8. The fitting process and prediction result of operational pressure under the first-order autocorrelation model. The circles denote the actual operational data; the solid line the verification fitting result using Eqs. (10) and (11); the dotted line the prediction using the calculated parameters: $\beta'_P = 1.398$ kg/cm^2, $\beta'_J = 5.578 \times 10^{-2}$ [kg/cm^2]/[m^2/s], $\beta'_C = 5.63 \times 10^{-4}$ [kg/cm^2]/[μS/cm], $\gamma' = 1.380 \times 10^{-1}$/°C, $\omega' = 4.273 \times 10^{-2}$/day and $\rho = 0.951$ with $R^2 = 0.950$. The parameters were estimated using data only between days 255 and 600, and the later prediction used the known information for independent variables such as C, T, and J with respect to time t.

nentially changes with temperature, the right hand side of Eq. (9) has the linear term of the temperature. In addition, the exponential increase of the pressure with respect to time is assumed to represent the average percent-change rate of the pressure between two consecutive cleaning operations.

Fig. 7 compares the semi-loglinear fitting and actual observational data. The semi-loglinear model of Eq. (9) captures the trend of the pressure profile, although the insensitiveness of the logarithmic function lead to deviation of the model from the observation: $R^2 = 0.619$. The value of β_J is at least one order of magnitude larger than β_C, which verifies that the pressure is much more influenced by the permeate flux than the feed concentration. This is because the low feed conductivity does not generate significant osmotic pressure. The positive value of γ indicates that the temperature and the pressure are anti-correlated. In other words, an increase in temperature leads to viscosity decline so that less pressure is required to produce the same flux. The value of $\omega = 2.03 \times 10^{-3}$ day$^{-1} \simeq 1.4 \times 10^{-2}$ week^{-1} implies that in order to maintain the predetermined flux, the pressure needs to be increased about 1.4% every week between two cleaning events; moreover, ω can be alternatively considered as the membrane fouling rate.

4.2.2. Autocorrelation model

In this section, we apply the first order autocorrelation model, in which the pressure is assumed to be a linear function of J, C, T and t, and an error at time t is influenced by a previous error:

$$\Delta P(t) = \beta'_P + \beta'_J J(t) + \beta'_C C(t) - \gamma' T(t) + \omega' t + e(t) \quad (10)$$

$$e(t) = \rho e(t-1) + \nu(t) \quad (11)$$

where $\nu(t)$ is the uncorrelated random variable with zero mean and a constant variance, capturing all omitted factors affecting the pressure in the model. In comparison to the semi-loglinear model that provided only the asymptotic variation of pressure behavior, the pure linearity of this autocorrelation model can capture the sensitive variation of the pressure in response to other filtration variables.

We used only partial data from days 255 to 600 (arbitrarily chosen) to calculate the coefficients in Eqs. (10) and (11). The data set does not depend on the time interval as long as the interval contains enough cleaning events so that the pressure growth rate can be reasonably estimated. Fig. 8 contains the actual pressure measurements, the fitting regions for estimating parameters in Eqs. (10) and (11), and the forecasting results compared with the observational data after day 600. The autocorrelation model has a higher accuracy ($R^2 = 0.949$) than that of the semi-loglinear model. Direct comparison of coefficients does not provide meaningful engineering information because they are not dimensionless. However, the positiveness of all the coefficients indicates that the autocorrelation model captures the underlying filtration phenomenon with the model stationarity confirmed by $\rho = 0.949 < 1$ [30]. The value of $\omega' = 0.0427$ kg/(cm^2 day) = 0.3 kg/(cm^2 week) indicates a weekly growth rate of pressure while ω of the semi-loglinear model showed the weekly percent-increase rate. Using the average pressure value over the entire filtration duration, the average percent increase rate of the semi-loglinear model is roughly equivalent to 0.1 kg/(cm^2 week), which is in the same order of magnitude of ω'.

The autocorrelation model with estimated coefficients is used to forecast future variations in pressure (after day 600) as shown in Fig. 8. This forecasting uses time series of C, T and J after day 600 as independent inputs. As expected, this autocorrelation model almost perfectly captures future pressure variations. Need for the independent inputs fundamentally limits the general applicability of this autocorrelation model. If annual average values of C, J and T are used, then the model only provides a constant value, independent of time t, of the pressure over the future period. However, one can easily make reasonable forecasts of the pressure using monthly, seasonal or yearly averages of C, J and T, based on the data availability over the past filtration duration, or employ independent extrapolations of time series for C, J and T.

5. Conclusions

We investigated statistical characteristics of filtration data (provided by Woongjin Chemical) using econophysics methods and further estimated membrane performance evolution (as pressure growth rate) using econometrics models developed in this study. The stationarity of log-changes in concentration, temperature, flux and pressure allowed the use of equilibrium-based econophysics methods, which observe the multiple filtration variables at the same point in time. The filtration index tree comprehensively visualizes underlying inter-correlations and dynamic similarities between the filtration variables in the constant flux operation. Log-changes of the filtration variables, in general, follow the normal distribution, which indicates that the filtration variables themselves are governed by the log-normal distribution due to periodic

cleaning and/or seasonal variations. The periodicity of the pressure time series was investigated using the autoregressive model of econometrics. Weekly, monthly, and seasonal correlations of the pressure profile were observed, and, more importantly, the strongest correlation was found between two consecutive days, indicating the resemblance of the pressure time series to a Markov chain. The semi-loglinear model, developed on the basis of Darcy's law, calculated the pressure growth rate between two cleaning events as 1.4 % per week, equivalent to 0.1 kg/cm^2 per week. Insensitivity of the semi-loglinear model was supplemented using the first-order autocorrelation model that has a linear functional relationship between the filtration variables and provided the pressure growth rate, 0.3 kg/cm^2 per week. Membrane performance evolution in terms of the pressure growth rate, as related to the membrane fouling rate, was similarly estimated using the two econometrics models with the stationarity verified using econophysics tools.

To the best of our knowledge, this paper is the first to statistically analyze actual filtration data using fundamental tools developed by econophysics and econometrics. Further studies can include advanced methods such as Bayesian analysis to determine the probability that tomorrow's pressure will stay same as the today's measured pressure given 5 % of flux decline today; a (generalized) autoregressive conditional heteroskedasticity (ARCH) model to capture underlying dynamic mechanisms which can have sophisticated distributions with a number of variables; and forecasting models for evaluation of basic characteristic parameters that describe long-term (e.g., yearly or longer) membrane aging rates.

Acknowledgement

This research was supported by a grant from the US National Science Foundation Faculty Early Career (CAREER) Development Program (CBET04-49431) and the International Scholar Program of Kyung Hee University, Korea. The corresponding author would like to specially thank Prof. MinKyung Kim with Konkuk University for illuminating discussions.

Nomenclature

α_i	coefficient of time-lag i days in Eq. (8)
β_C	coefficient for log concentration in Eq. (9)
β'_C	coefficient for concentration in Eq. (10) [(kg/cm^2)/(μS/cm)]
β_J	coefficient for log flux in Eq. (9)
β'_J	coefficient for flux in Eq. (10) [(kg/cm^2)/(m^3/s)]
β_P	constant shift term in Eq. (9)
β'_P	constant shift term in Eq. (10) [kg/cm^2]
$\Delta\pi$	osmotic pressure gradient [kg/cm^2]
ΔP	applied pressure [kg/cm^2]
Δt	time interval [day]
ϵ	disturbance in Eq. (9)
γ	coefficient for temperature in Eq. (9) [1/K]
γ'	coefficient for temperature in Eq. (10) [1/K]
μ	feed water viscosity [Pa s]
$\nu(t)$	uncorrelated random variable with zero mean and a constant variance in Eq. (11)
ω	coefficient for time in Eq. (9) [1/day]
ω'	coefficient for time in Eq. (10) [1/day]
ρ_{ij}	correlation coefficient
$\tilde{S}_i(t)$	normalized $S_i(t)$
C	feed concentration measured as conductivity [μS/cm]
d_{ij}	Euclidean distance between i and j
e	disturbance in Eq. (11)
J	permeate flux measured as flow rate [m^3/s]
P_0	constant swift term of autoregressive model in Eq. (8) [kg/cm^2]
$P_i(t)$	daily closure stock price of company i [$]
$R_i(t)$	return (or growth rate) of daily closure stock price of company i
R_m	membrane resistance [1/m]
$S_i(t)$	logarithmic change of daily closure stock price of company i, defined in Eq. (11)
T	temperature [K]

References

[1] E. Zondervan, B. Roffel, Modeling and optimization of membrane lifetime in dead-end ultrafiltration, J. Membr. Sci. 322 (1) (2008) 46–51.
[2] H. Niemi, A. Bulsari, S. Palosaari, Simulation of membrane separation by neural networks, J. Membr. Sci. 102 (1995) 185–191.
[3] M. Razavi, A. Mortazavi, M. Mousavi, Dynamic modelling of milk ultrafiltration by artificial neural network, J. Membr. Sci. 220 (1–2) (2003) 47–58.
[4] M. Hamed, M. Khalafallah, E. Hassanien, Prediction of wastewater treatment plant performance using artificial neural networks, Environ. Modell. Softw. 19 (10) (2004) 919–928.
[5] H. Chen, A.S. Kim, Prediction of permeate flux decline in crossflow membrane filtration of colloidal suspension: a radial basis function neural network approach, Desalination 192 (1–3) (2006) 415–428.
[6] V. Pareto, Cours d'Economie Politique, Lausanne and Paris, 1897.
[7] P. Levy, Calcul des Probabilities, Gauthier-Villard, Paris, 1925.
[8] B. Mandelbrot, The variation of certain speculative prices, J. Busi. 36 (4) (1963) 394–419.
[9] H.E. Stanley, Introduction of Phase Transitions and Critical Phenomena, Oxford University Press, Oxford, 1971.
[10] R. Brown, A brief account of microscopical observations, Phil. J. (1828) 358–371.
[11] L. Bachelier, Theorie de la speculation, Annales Scientifiques de l'Ecole Normale superieure 17 (1900) 21–86.
[12] A. Einstein, Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen, Ann. Phys. 322 (8) (1905) 549–560.
[13] N. Wiener, Differential space, J. Math. Phys. 58 (1923) 31–174.
[14] C. Gardiner, Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences, Springer, 1996.
[15] F. Black, M. Scholes, The pricing of options and corporate liabilities, J. Pol. Econ. 81 (3) (1973) 637–654.
[16] M. Stanley, L. Amaral, S. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. Salinger, H. Stanley, Scaling behaviour in the growth of companies, Nature 379 (6568) (1996) 804–806.
[17] H. Stanley, V. Afanasyev, L. Amaral, S. Buldyrev, A. Goldberger, S. Havlin, H. Leschhorn, P. Maass, R. Mantegna, C. Peng, P. Prince, M. Salinger, M. Stanley, G. Viswanathan, Anomalous fluctuations in the dynamics of complex systems: from DNA and physiology to econophysics, Physica A 224 (1–2) (1996) 302–321.
[18] H. Stanley, L. Amaral, D. Canning, P. Gopikrishnan, Y. Lee, Y. Liu, Econophysics: can physicists contribute to the science of economics? Physica A 269 (1) (1999) 156–169.
[19] R. Mantegna, Z. Palágyi, H. Stanley, Applications of statistical mechanics to finance, Physica A 274 (1–2) (1999) 216–221.
[20] X. Gabaix, P. Gopikrishnan, V. Plerou, E. Stanley, A unified econophysics explanation for the power-law exponents of stock market activity, Physica A 382 (1) (2007) 81–88.
[21] A. Arnéodo, J.-F. Muzy, D. Sornette, Direct causal cascade in the stock market, Eur. Phys. J. B 2 (2) (1998) 277–282.
[22] P. Cizeau, Y. Liu, M. Meyer, C. Peng, H.E. Stanley, Volatility distribution in the S&P500 stock index, Physica A 245 (3–4) (1997) 441–445.
[23] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Turbulent cascades in foreign exchange markets, Nature 381 (6585) (1996) 767–770.
[24] W. Li, Absence of 1/f spectra in Dow Jones daily average, Int. J. Bifurcat. Chaos 1 (3) (1991) 583–597.
[25] L. Amaral, S. Buldyrev, S. Havlin, M. Salinger, H. Stanley, Power law scaling for a system of interacting units with complex internal structure, Phys. Rev. Lett. 80 (7) (1998) 1385–1388.
[26] N. Vandewalle, M. Ausloos, Coherent and random sequences in financial fluctuations, Physica A 246 (3–4) (1997) 454–459.
[27] E. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (4) (1957) 620.
[28] E. Jaynes, Information theory and statistical mechanics. II, Phys. Rev. 108 (2) (1957) 171.

[29] M. Bayes, M. Price, An essay towards solving a problem in the doctrine of chances, Phil. Trans. Roy. Soc. London 53 (1763) 370–418.
[30] P. Kennedy, A Guide to Econometrics, Balckwell Publishing, 2008.
[31] A. Giffin, Maximum entropy: the universal method for inference, Ph.D. Thesis, Department of Physics, the University at Albany, State University of New York, 2009.
[32] A. Giffin, From physics to economics: an econometric example using maximum relative entropy, Physica A: Stat. Mech. Appl. 388 (8) (2009) 1610–1620.
[33] R. Frisch, Correlation and scatter in statistical variables, Nordic Stat. J. (1929) 36–102.
[34] R.N. Mantegna, H.E. Stanley, An Introduction of Econophysics, Cambridge University Press, 2000.
[35] J.W. Gibbs, Elementary Principles in Statistical Mechanics, Charles Scribner's Sons, New York, 1902.
[36] E. Jaynes, Gibbs vs Boltzmann entropies, Am. J. Phys. 33 (5) (1965) 391–398.
[37] S. Kullback, Information Theory and Statistics, Dover Publications, 1997.
[38] R.N. Mantegan, Degree of correlation inside a financial market, in: Applied Nonlinear Dynamics and Stochastic Systems near the Millennium, AIP Press, New York, 1997, pp. 197–202.
[39] D.B. West, Introduction to Graph Theory, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[40] R. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1) (1999) 193–197.
[41] R. Rammal, G. Toulouse, M. Virasoro, Ultrametricity for physicists, Rev. Mod. Phys. 58 (3) (1986) 765–786.
[42] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, M. Virasoro, Nature of the spin-glass phase, Phys. Rev. Lett. 52 (13) (1984) 1156–1159.
[43] C.H. Papadimitriou, K. Steigitz, Combinatorial Optimization, Prentice-Hall, Englewood Cliffs, NJ, 1982.
[44] F. Reif, Fundamentals of Statistical and Thermal Physics, McGraw-Hill, New York, 1965.
[45] R. Blattberg, N. Gonedes, A comparison of the stable and student distributions as statistical models for stock prices, J. Busi. 47 (2) (1974) 244–280.
[46] P. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41 (1) (1973) 135–155.
[47] R. Mantegna, H. Stanley, Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight, Phys. Rev. Lett. 73 (22) (1994) 2946–2949.
[48] R. Blandford, D. Eichler, Particle acceleration at astrophysical shocks: a theory of cosmic ray origin, Phys. Rep. 154 (1) (1987) 1–75.
[49] S. Chandrasekhar, Brownian motion, dynamical friction, and stellar dynamics, Rev. Mod. Phys. 21 (3) (1949) 383–388.