

# ENTROPY AS A MEASUREMENT FOR THE QUALITY OF DEMAND FORECASTING

---

Bernd Scholz-Reiter, Jan Topi Tervo, Uwe Hinrichs

*Department of Planning and Control of Production Systems, University of Bremen  
bsr, ter, hin@biba.uni-bremen.de*

*Production planning and control is a highly complex process influenced by many factors. An important part of this broad task is demand forecasting, for which many methods already have been developed. But due to the occurring dynamics in the used data, the prediction may differ strongly from the optimum and thus errors leading to rising costs are inevitable. In this paper we will propose the entropy as a measurement for the quality of demand forecasting respectively as relative estimation for the forecasting error. In general, entropy is a measurement for disorder and thus also for information content. Since lack of information leads to inaccuracy of forecasting, the entropy can be identified with the quality of demand prediction. First results on the basis of time-series obtained from mathematical functions, discrete-event simulations of a production network scenario and a real shop-floor system will show the successful transfer of this method.*

## 1. INTRODUCTION

Nowadays, production planning and control is a challenging task due to changing market conditions and increasing dynamics in global network organizations. Its primary objective is to schedule and realize the ongoing production plan efficiently (Eversheim et al., 1996). To do so, the production capacities and the needed amount of resources have to be regarded.

While the number of machines is constant in general, the demand for any kind of resource or material has to be forecasted and ordered with respect to the planned output in a defined time-period. Incorrect or invalid forecasts can lead to severe consequences: ordering too much material will result in higher stocks with rising costs for stock-holding and materials. When ordering less than the needed amount the risk of production downtimes arises. Therefore, exact and secure methods for the important process of demand forecasting are needed.

This becomes clearer when looking at the several factors or sources of information which influence the planned demand for materials in a defined period of time. First of all, exact numbers from the sales market are needed, which determine the production plan. Here, seasonal fluctuations can occur depending on the kind of produced goods. Furthermore, increasing dynamics in present markets have been observed and nonlinear effects in production systems or production networks have been verified (Scholz-Reiter et al., 2003), (Wiendahl et al., 2000). Regarding all

these factors and the possible economic results, it is obvious that adaptive and trustful methods have to be used when forecasting the demand.

Until now several approaches for demand forecasting based on statistical and mathematical techniques are used. The future demand is forecasted by using a time-series consisting of former values (Granger, 1989). Although these methods were tested and show a strong reliability, there are still no means to measure the quality of the calculated result. But since demand forecasting techniques basically depend on preceding information, a measurement for the prediction quality should be based on the available information content. Hence, we propose the entropy to characterize the quality of demand forecasting respectively the relative estimation for the forecasting error, as entropy is a measurement for disorder and, thus, also for information content.

In the following, several forecasting methods as well as the entropy in general are presented. The next step is to apply these techniques to different time-series showing demand values and compare the measured forecasting error with the calculated entropy. Therefore, several time-series taken from different mathematical functions, discrete-event simulations of a production network scenario and a real shop-floor system were used.

## 2. FORECASTING METHODS

In recent publications many forecasting methods have been proposed for a broad variation of different settings (Makridakis et al., 1998). But since the focus of this paper is rather on the quality of the forecasting than on the method itself, we will concentrate on two different basic techniques, which will be presented briefly here.

The first is the moving average approach, which is best suitable for simple time-series with identifiable fluctuations around a mean value and without cycles. Thereby, not all available data from a time-series is used, but only the last  $n$  values. The demand in the future period  $\lambda_{i+j}$  is then determined by the averaged demand  $\lambda_i(n)$  of the considered past period (Granger, 1989):

$$\lambda_{i+1} = \lambda_i(n) = \frac{1}{n} \sum_{j=i-n}^i \lambda_j. \quad (1)$$

The number of values given by  $n$  allows looking at a limited time segment and thereby shows a high flexibility. But also,  $n$  influences the reaction to changes: a large value of  $n$  neglects rapid changes, while a smaller value of  $n$  follows fast dynamics.

The other applied method is exponential smoothing or exponentially weighted moving average (Granger, 1989). Here the future demand  $\lambda_{i+j}$  is calculated from the weighted average of the measured demand  $\lambda_i$  and the forecasted demand  $\lambda_i(\alpha)$  of the past period:

$$\lambda_{i+1}(\alpha) = \alpha \lambda_i + (1 - \alpha) \lambda_i(\alpha) \quad \text{with } 0 < \alpha < 1. \quad (2)$$

When including the  $n$  past values, this leads to the weighted average of the data:

$$\lambda_{i+1}(\alpha) = \alpha \sum_{j=i-n}^i (1 - \alpha)^{i-j} \lambda_j. \quad (3)$$

The factor  $(1 - \alpha)^{i-j}$  causes an exponential decrease of the influence of the past values on the average. If  $\alpha$  is near one, the decay is strong, i.e., the effect of the past

values is weak. In contrast, when  $\alpha$  is near zero, the decay is weaker and past values are taken into account more strongly. The challenge is to find a suitable value for  $\alpha$ . Until now there is no objective way to define this factor.

### 3. ENTROPY

Commonly, the word entropy is associated with disorder, uncertainty or ignorance. It originates from two different domains of science, namely physics and information theory. Both derivations have similarities, but require knowledge in each domain.

Entropy as a measure with physical meaning was introduced by Clausius (1865) and later precised by Boltzmann (1880). In thermodynamics, a macroscopic state is described by the microscopic behaviour of all its  $N$  particles. These are defined by their positions and their impulses, which span a  $6N$ -dimensional phase space. Entropy then gives a measurement for the quantity of different possible microstates of that thermodynamical system or the volume of phase space occupied by it. In other words, it describes the internal disorder within a system. Since entropy in statistical physics gives a probabilistic treatment to a system's thermal fluctuations, higher entropy also means a greater lack of information on the exact configuration of the system. Hence, it has many similarities with entropy derived in information theory. This definition is principally based on Shannon (Shannon, 1948) and in this sense it is a measure for the amount of randomness hidden in an information sequence. This means that a sequence with redundancies or statistical regularities exhibits small values of entropy and in contrast, a uniform distribution of sequence symbols, e.g., white noise, leads to the highest entropy value. As a consequence, history and future of that sequence are completely uncorrelated. Since this paper is focussed on time-series analysis, the information theoretical definition of entropy will be considered.

#### 3.1 Symbolic dynamics

In order to calculate a value for entropy, a sequence of symbols is needed. In time series with discrete values, e.g., buffer levels, this condition is granted. But for a continuous variable, the values have to be transformed into an adequate sequence. Nevertheless, the number of discrete values may also be reduced by such a transformation. In physics, this method is well known as 'symbolic dynamics' (Hao, 1988).

Apparently, when transforming a time-series into a symbol sequence, a large amount of detailed information is lost, but some invariant and robust characteristics such as periodicity, symmetry and chaos can be conserved. But this strongly depends on the choice of the transformation. Due to the reduction of details, the analyses of symbol sequences are less vulnerable to noise (Daw et al., 2003) and, consequently, conclusions drawn from these sequences are more precise.

Before calculating the entropy, the first choice to make is the alphabet size  $|A|=l$ , i.e., the number of symbols used to transform the original time-series into a sequence of symbols. This variable determines how much of the original information is conserved. The simplest case is a binary alphabet with  $l=2$  and  $A=\{0,1\}$ . The next step is to decide about the transformation itself. There are two

elementary different ways: static and dynamic transformations (for illustration see Figure 1). The static transformation is realized by choosing one (in the binary case) or more thresholds and the different symbols are then assigned to the intervals between them. There are diverse rules to calculate those thresholds, e.g., data mean or median value (Daw et al., 2003). The dynamic transformation is preferred when the dynamics are more important than the absolute values. Thereby, step-to-step differences in the sequence are taken into account and, in the binary case for example, a positive difference leads to one and a negative to the other symbol. Of course it is possible to make a bad choice for the transformation, hence all or at least the relevant information is lost.

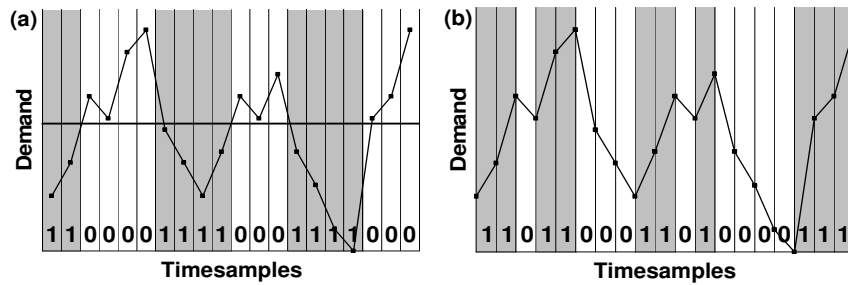


Figure 1 - Illustration of symbol sequence generation. In (a) a binary static and in (b) a binary dynamical transformation is shown.

### 3.2 Calculation of entropy

In order to calculate the Shannon-Entropy symbol sequence statistics have to be performed. More precisely, a histogram of repeating sequences of length  $L$  has to be obtained. Therefore,  $L$  consecutively following symbols  $s$  are combined to a word  $s^L$  and every word is uniquely coded to a decimal number (see Figure 2) to avoid the handling of long symbol sequences. Figuratively, one can think of a window of size  $L$  being slid from the beginning to the end of the sequence and at every position a word of length  $L$  is found. Then, a histogram of the relative word frequencies  $p(s^L)$  can be obtained and with

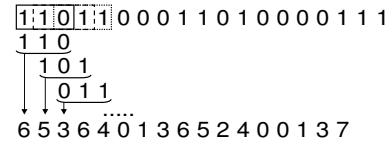


Figure 2: The code sequence (lower row) is produced by a window of length  $L=3$  being slid over the symbol sequence (upper row).

$$H_s = - \sum_{s^L \in A^L} p(s^L) \log_2 p(s^L) \tag{4}$$

the Shannon-Entropy can be calculated as the sum over all possible words of length  $L$ . Since the value of the entropy is strongly dependent on the word length  $L$ , a standardisation with the maximally possible entropy is required:

$$H = \frac{H_s}{H_{MAX}} \in [0,1] \tag{5}$$

This maximum value is obtained for a uniform distribution of the word frequencies

$$p(s^L) = \frac{1}{L^L}, \forall s^L \quad (6)$$

and thus  $H_{MAX}=L$ . This leads to a zero entropy for constant sequences of symbols and to  $H=1$  for a completely random symbol sequence.

#### 4. MEASUREMENT FOR QUALITY OF DEMAND FORECASTING

The entropy as a reliable measurement for demand forecasting quality is evaluated by comparing the forecasted demand with the real demand value of the next time step. Then the correlation between forecasting error and calculated entropy is identified. But previously, the parameters  $l$  (alphabet size),  $L$  (word length),  $n$  (time horizon) and  $\alpha$  (smoothing factor) in Equations (1)-(5) have to be determined.

The maximum possible word length strongly depends on the alphabet size and the length of the time-series: The larger the alphabet and the shorter the time-series, the smaller is the possible word length (and vice versa) (Daw et al., 2003). Since real time-series are in general rather short, we use a binary alphabet to be able to get a word length up to  $L=5$ . The data mean was used as a threshold for a static transformation, except for the real shop-floor system where additionally a dynamical transformation was applied. Also, for the shop-floor system it was unfortunately only possible to calculate the entropy up to a word length  $L=4$  due to the shortness of the time-series of only 360 samples. To be as realistic as possible, the time horizon is chosen to be half a year, i.e., with one time sample per day. This leads to  $n=150$  time samples and additionally, the smoothing parameter was found to produce in average best results for  $\alpha=0.76$ .

##### 4.1 Simple examples

To depict the properties of the entropy, its value for three simple time-series generated from a constant, a sine and a uniformly distributed random function (see Figure 3) is used. The calculated values are enlisted in Table 1. As stated above, a constant function leads to a single peak in the distribution of generated words and thus zero entropy will follow. The computational calculation for different word lengths confirms this result. As a consequence,

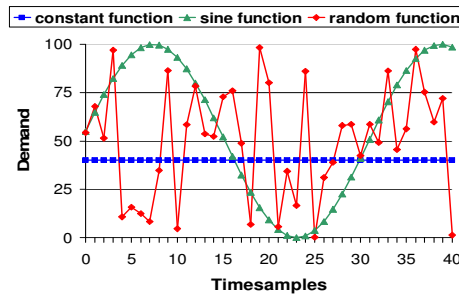


Figure 3: Extract from the constant, sinusoidal and random function, respectively the generated time-series (denoted as points), used for entropy calculation and forecasting.

forecasting without error is possible. Contrarily, a time-series of random values leads to a uniform distribution of generated words and hence to a maximum entropy of value one; an exact forecasting is impossible. The sine function produces for increasing word lengths decreasing entropy values (see. Table 1) with a mean of approximately 0.5, because a longer word implies more information and a better predictability.

The forecasting of a constant time-series is trivial. Both methods (gliding average and exponential smoothing) will deliver exact results of future demand without any error. This coincides with the entropy value of zero. Similarly, the forecasting error for the random function corresponds to the calculated entropy. Here, the forecasting method has no significance, since the past values do not correlate at all with future values. This is shown by the forecasted values (calculated with gliding average method) varying only a little around 50 and their error to the real demand, fluctuating between 0 and 100. Table 3 enlists these values for three randomly picked points in time of the time-series. The prediction of the demand for the sine function is in average better with the exponential smoothing than with the gliding average method.

For the forecasting error it is of major importance at which time step a prediction is made. Around the minimum and maximum values of the function good values can be obtained, while rather large errors occur when the slope is large. This reflects the entropy value of about 0.5 calculated in Table 1.

#### 4.2 Simulation time-series

To generate more application-oriented time-series a discrete event simulation model of a supply chain of four enterprises with external customer driving (Scholz-Reiter et al., 2005) was used. The customer demand was realized by a discrete sinusoidal and a uniformly distributed random function (see Figure 4). The entropy values calculated for both time-series are comparable to those calculated in Section 4.1. The random demand leads to a random fluctuation in the time-series and so an entropy value of one follows. On the other hand, the sinusoidal demand causes a deterministic structure similar to the sine function and accordingly an entropy value

Table 1: Entropy values of different word lengths for the three different time-series generated by a constant, sine and random function.

Word length	Entropy		
	constant	sine	random
3	0.00	0.56	1.00
4	0.00	0.50	1.00
5	0.00	0.47	1.00

Table 2: Entropy values of different word lengths for the two different time-series generated by the DES model with a sinusoidal and random customer demand and entropy values of different word lengths for the time-series of the real shop-floor system created with static and dynamical transformation.

Word length	Entropy			
	Simulation Data		Real Data	
	Sine	Random	Dynamical	Static
2	-	-	0.95	0.90
3	0.61	1.00	0.93	0.90
4	0.56	1.00	0.92	0.90
5	0.52	1.00	-	-

of about 0.57 is calculated (see. Table 2). These values correspond to the forecasting errors. As shown in Table 3 the sine function can be forecasted well because the used forecasting methods deliver best results when only marginal dynamics are present. Again a comparison for randomly picked points in time of the time-series were done (see Table 3).

### 4.3 Real data

The last time-series to be analysed is taken from the demand of a real shop-floor system (see. Figure 5). Here, the entropy is calculated with static and dynamical transformations to constitute the differences between them for this time-series. As shown in Table 2 the values differ only slightly with values of 0.9 and 0.93 respectively. Again, the entropy value corresponds to the predictability of the time-series: It shows an almost random behaviour with only little determinism. This is confirmed by the calculated forecasting values compared with real demand (see Table 3).

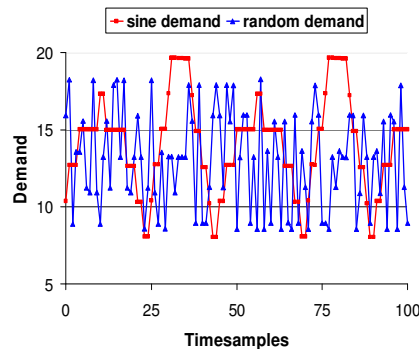


Figure 4: Extract from two different time-series generated by the DES model with a sinusoidal and random customer.

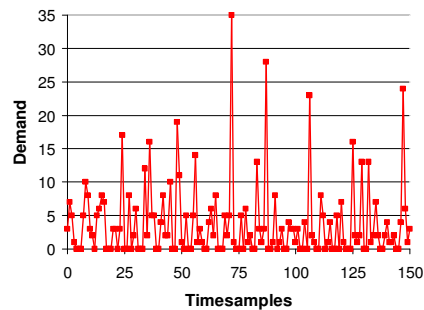


Figure 5: Extract from the demand time-series of a real shop-floor system.

## 5. SUMMARY

The entropy can be calculated quickly and easily for rather short word length (up to  $L=8$ ) and realistic demand time-series of a length of max. 10000 time steps. Since it is a measurement for uncertainty it corresponds to the predictability of time-series. Therefore, no absolute forecasting error can be obtained, but a graduation between 0 (perfectly predictable) and 1 (not predictable at all) is very well possible.

The presented results show that this property of the entropy can be successfully transferred to relatively measure the reliability of demand forecasting. But for a promising application in order forecasting methods further research has to be done, which will deal with the evaluation of the several parameters and concrete recommended actions especially.

Table 3: Forecasted and real demand values for all mentioned time-series for randomly picked points in time, respectively their relative error.

Time-Series	Forecasted value	Real Value	Relative Error
sine function	29.1	18.7	55.6%
	84.9	92.4	8.1%
	1.6	1.4	14.3%
random function	53.9	60.9	11.5%
	53.3	77.9	31.6%
	53.1	18.5	187.0%
sinusoidal demand	15.1	12.59	19.9%
	15.03	14.99	0.3%
	10.88	10.31	5.5%
random demand	13.33	18.38	27.5%
	13.55	10.85	24.9%
	13.41	11.35	15.4%
Real data	3.01	3	24.0%
	3.05	1	205.0%
	3.42	11	68.9%

## 6. ACKNOWLEDGEMENTS

This work is funded by the German Research Foundation (DFG) under the reference number Scho 540/13-1 "Synchronisation of the nodes in production and logistics networks" and the Volkswagen Foundation under the reference number I/78 217 "Modelling and Analysis of Production and Logistics Networks Using Methods of Nonlinear Dynamics"

## 7. REFERENCES

1. Daw, C. S., Finney, C. E. A., Tracy, E. R.. A review of symbolic analysis of experimental data. *Review of Scientific Instruments* 74, 2003, pp. 915-930.
2. Eversheim, W., Schuh, G.. *Produktion und Management Bd. 2*. Springer, 1996.
3. Granger, C. W. J.. *Forecasting in Business and Economics*. Academic Press, London, 1989.
4. Hao, B.-L.. *Elementary Symbolic Dynamics and Chaos in Dissipative Systems*. World Scientific, 1988.
5. Makridakis, S., Wheelwright, S. C., Hyndman, R. J.. *Forecasting*. Wiley and Sons, 1998.
6. Scholz-Reiter, B., Freitag, M.. On the Dynamics of Manufacturing Systems – A State Space Perspective. *Proceedings of the 36th CIRP-International Seminar on Manufacturing Systems*, 2003, pp. 455-462.
7. Scholz-Reiter, B., Hinrichs, U., Delhoum, S.. Analyse auftretender Instabilitäten in dynamischen Produktions- und Logistiknetzwerken. In: *Industrie Management* 21 (2005) 5, S. 25-28.
8. Shannon, C. E.. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948, pp. 379-423 and pp. 623-656.
9. Wiendahl, H.-P., Worbs, J.. Simulation based analysis of complex production systems with methods of nonlinear dynamics. *IMCC'2000 International Manufacturing Conference in China*, 2000.