J. Japan Statist. Soc.
Vol. 38 No. 1 2008 119–130

CONTRIBUTIONS OF PROFESSOR HIROTUGU AKAIKE IN STATISTICAL SCIENCE

Genshiro Kitagawa*

Prof. Akaike made significant contributions in various fields of statistical science, in particular, in time series analysis in frequency domain and time domain, information criterion and Bayes modeling. In this article, his research contributions are described in order of launching period, frequency time domain analysis, time domain time series modeling, AIC and statistical modeling, and Bayes modeling.

Key words and phrases: ABIC, AIC, Bayes modeling, FPE, frequency domain analysis, time series analysis, time series models, TIMSAC.

1. Introduction

Professor Hirotugu Akaike was awarded the 2006 Kyoto Prize for "his major contribution to statistical science and modeling with the Akaike Information Criterion (AIC)". In 1973, he proposed the AIC as a natural extension of the log-likelihood. The most natural way of applying the AIC is to use it as the model selection or order selection criterion. In the MAICE (minimum AIC estimation) procedure, the model with the minimum value of the AIC is selected as the best one among many possible models. This provided a versatile procedure for statistical modeling that is free from the ambiguities inherent in application of the hypothesis test procedure.

However, the impact of the AIC is not limited to the realization of an automatic model selection procedure, and it eventually led to a paradigm shift in statistical science. In conventional statistical inference, the theories of estimation and test are developed under the assumption of the presence of a true model. However, in statistical modeling, the model should be constructed based on the entire knowledge such as the established theory, empirical facts, current observations and even the objective of the analysis. Prof. Akaike gave a practical answer to the selection of the prior distribution of the Bayes model. Due to the development of information technologies, we can now access to huge amounts of data in various fields of science and social life. In this information and knowledge society, the Bayes model is becoming a key technology.

In this article, we shall look back at his research in five stages, namely, the launching period, frequency domain time series analysis, time series modeling, AIC and statistical modeling and Bayes modeling (Parzen *et al.* (1998)). It should be noted here that the reader will notice that his research was always performed based on the needs of researchers in the real-world.

Accepted November 20, 2007.

*The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan.

2. Launching period: Making contacts with engineers

In 1952, Professor Akaike graduated from the Department of Mathematics, The University of Tokyo and became a researcher at the Institute of Statistical Mathematics. During the 1950's, he published several theoretical papers related to decision processes, evaluation of probability distributions, computation of eigenvalues, and the Monte Carlo method for solving linear equations. Among these works, the most famous one is the convergence analysis of the optimum gradient method (Akaike (1959a)). In this paper, he analyzed the limiting behavior of a probability distribution when a type of transformation was repeatedly applied to an initial distribution, and showed the convergence property of the optimum gradient method. This result became a foundation in the development of more sophisticated nonlinear optimization methods and is introduced in a standard textbook of nonlinear optimization methods (Kowarik and Osborne (1968)).

During this launching period, however, he was rather interested in real-world problems and tried to develop substantial contacts with engineers in various fields of industries. Through these contacts, he realized that the conventional data analysis methods often are not able to yield interesting results for real-world problems, and that we should develop a model that fully takes into account the structure of the process. In the analysis of traffic density on a road, he considered a zero-one process and, under the assumption of the independence of the lengths of time intervals between cars, he derived the gap process for the structure of the series (Akaike (1956)). Then, in collaboration with Dr. Shimazaki of the Sericultural Experiment Station, Ministry of Agriculture, he developed a control method for silk filature production processes, based on gap process modeling (Akaike (1959b)). This method provided a reference process that could be used for the detection of abnormalities in the actual reeling process. By extending this method, Dr. Shimazaki brought significant innovation to the method of silk production in Japan.

It should be emphasized here that from a very early stage in the 1950's, he realized the importance and necessity of modeling the structure of an object.

3. Frequency domain time series analysis

In the early 1960's, however, he suddenly changed his research policy. During his communications with industrial engineers, he became to understand that linear stationary models were used frequently and that they had a very wide range of applications. Further, he noticed that there were many unsolved important problems that statisticians should contribute to. He therefore, changed his mind and tried to develop practical analysis methods based on linear stationary models. Through collaboration with Dr. Kaneshige of the Isuzu Motor Company, he learned the Blackman-Tukey method, and realized that a practical method of spectrum estimation can be developed by using a smoothing method. In this area of spectrum smoothing, he proposed a spectral window with negative coefficients so that the peak of the spectrum is not reduced significantly (Akaike

(1962)). He suggested comparing the results of smoothing by positive windows and non-positive windows.

Then, by the cooperative research with Dr. Yamanouchi of the Transportation Technical Research Institute, he developed a practical method of estimating the frequency response function by solving the problem of phase-shift by shifting the time axis of the cross-covariance function (Akaike and Yamanouchi (1962)). This method made it possible to estimate the frequency response function from observations under normal steering without using sinusoidal inputs to systems with various frequencies. He then organized a workshop and tried to apply the method to various areas of engineering such as, vibrations, engines of cars, the roll of a ship, the response of an airplane to cross winds, hydroelectric power plants, underground structures based on micro-tremors, tsunamis, the back rush of a nonlinear system, etc. The outcome of the workshop was reported in the supplement of AISM (Annals of the Institute of Statistical Mathematics) (Akaike (1964)).

However, during the joint work with these engineers, an important problem that cannot be solved by the conventional spectral analysis arose. Namely, most of the systems in the real-world such as ships, cars, airplanes, chemical plants, and economics contain some kind of feedback. In these feedback systems, an output from a subsystem A becomes an input to another subsystem B, and the output of the subsystem B becomes an input to the subsystem A. Then it is quite difficult to identify the real cause of the fluctuation in the feedback system. In particular, it cannot be solved by the conventional frequency domain approach, since in the frequency domain analysis; we cannot explicitly utilize the physical realisability of the system that imperatively exists in the real world physical system. In Akaike (1967), he pointed out this difficulty and the limitation of the frequency domain analysis.

4. Time domain time series modeling

At this opportunity, he returned to the time domain analysis. At that time, Chichibu Cement Company was preparing to apply optimal control theory to the control of a rotary kiln in a cement plant instead of the conventional PID controller. Conventional spectral analysis methods were useless for the analysis of a typical feedback system consisting of the raw material feed, fuel feed, gas damper angle, and temperature at various locations, etc. As stated in the previous section, this is because in the spectral analysis, we cannot properly take into account the restrictions necessary for physical realisability.

However, if we introduce a structural model that expresses the relation between variables $y_{n1}, \dots, y_{n\ell}$ with feedback,

$$y_{ni} = \sum_{j \neq i} \sum_{k=1}^{m} \alpha_{kij} y_{n-k,j} + \varepsilon_{ni}, \tag{4.1}$$

where m is the maximum time lag and ε_{ni} is a Gaussian white noise, then it is possible to perform causal analysis of a feedback system. Akaike (1968) used the fact that the above structural model for a feedback system can be expressed in a multivariate autoregressive model,

$$\boldsymbol{y}_n = \sum_{j=1}^{m} A_j \boldsymbol{y}_{n-j} + \boldsymbol{w}_n, \tag{4.2}$$

where $\boldsymbol{y}_n = (y_{n1}, \ldots, y_{n\ell})^T$, $\boldsymbol{w}_n = (\varepsilon_{n1}, \ldots, \varepsilon_{n\ell})^T$ and A_j is an $\ell \times \ell$ AR coefficient matrix, therefore the above structural model can be estimated from the time series data. In the paper, based on the identified model and the relation between the multivariate autoregressive model and the cross spectrum, the power contribution was defined by

$$r_{ij}(f) = \frac{|b_{ij}|^2 \sigma_j^2}{p_{ii}(f)}, \tag{4.3}$$

where $p_{ii}(f)$ is the power spectrum of the channel i at frequency f obtained by

$$P_{ii}(f) = \sum_{j=1}^{\ell} |b_{ij}|^2 \sigma_j^2, \tag{4.4}$$

and $B(f) = (b_{jk}) = (\sum_{j=1}^{m} a_j(j,k) \exp(-2\pi i j f))^{-1}$. The power contribution $r_{ij}(f)$ expresses the proportion of the contribution of noise of the variable y_j in the power of the fluctuation of the variable y_i at frequency f in the feedback system. This method was used to analyze not only the cement rotary kiln, but also in the analysis of nuclear and thermal power plants, living bodies, economic and financial systems, etc. (Akaike and Nakagawa (1988), Akaike and Kitagawa (1998)).

Thus, by the use of a multivariate autoregressive model, the analysis of feedback systems became practical. However, there was a difficulty with this method. It is not so difficult to estimate the parameters of a given autoregressive model. However, if we use an autoregressive model with different order, we will have quite different results in prediction or in the analysis. Therefore, if we do not have a reasonable and practical procedure to determine the order of autoregression, the method of analysis of a feedback system through the use of a time series model will not become practical.

To solve the problem, Akaike (1969a) proposed a predictive point of view. In conventional statistical analysis, we aim at reproducing the “true” structure as precisely as possible. On the other hand, he considered the situation where we use the estimated AR model for the prediction of future values that will be generated from the same structure as the one used for the estimation of the model. There was a significant difference between the modeling for the purpose of the estimation of the true structure and for the prediction of future values, and the introduction of the predictive points of view had a strong impact on statistical modeling thereafter.

In the simplest case of a univariate autoregressive model,

$$y_n = \sum_{j=1}^{m} a_j y_{n-j} + \varepsilon_n, \tag{4.5}$$

where m is the order and ε_n is a Gaussian white noise with mean 0 and variance σ^2, based on the predictive point of view, Akaike (1969a) proposed the final prediction error (FPE) criterion

$$\mathrm{FPE}_m = \frac{n+m+1}{n-m-1}\hat{\sigma}_m^2, \tag{4.6}$$

where n and m are the data length and the order of autoregression, respectively, and $\hat{\sigma}_m^2$ is the least squares estimate of the innovation variance of the autoregressive model with order m. With this criterion, the best model can be automatically determined by selecting the order that attains the smallest value of FPE. Since the power spectrum of the autoregressive model is given by

$$p(f) = \frac{\hat{\sigma}^2}{|1 - \sum_{j=1}^{m} \hat{a}_j \exp(-2\pi i j f)|^2}, \tag{4.7}$$

where $\hat{a}_j$ and $\hat{\sigma}_m^2$ are the estimated AR coefficient and innovation variance, respectively, an estimate of the power spectrum is automatically obtained by fitting an AR model (Akaike (1969b)). This AR method of spectrum estimation is equivalent to the MEM spectrum proposed by Burg (1968).

A multivariate version of FPE, MFPE, was also developed. However, the way of expressing the accuracy of the multivariate time series in a scalar criterion is not unique. The MFPE proposed in Akaike (1970) uses $\det \hat{\Sigma}$, where $\hat{\Sigma}$ is the determinant of the estimated variance-covariance matrix of the innovation. It is equivalent to use the likelihood function for the estimation of the parameters of the model.

As an important application of the multivariate AR model, Akaike (1971) proposed a statistical method of optimal control. Optimal control theory was established in the early 1960's. However, there was difficulty in applying it to large-scale complex systems or a system with large noise disturbances, since for such systems, it was difficult to derive the system model that expressed the dynamic behavior of the system. The difficulty in obtaining the system model was the bottleneck in applying optimal control theory in real-world problems. The statistical control method advocated by Prof. Akaike solved this problem by using the state-space representation of the multivariate autoregressive model. This method was first applied to the optimal control of a cement rotary kiln (Otomo *et al.* (1972)). In ship controlling, Otsu *et al.* (1979) applied the method for designing an autopilot system for improved course keeping navigation. Later, Mitsui Ship Building Company developed an autopilot system that reduces the rolling of ships in addition to keeping the course. This was anticipated by a power contribution analysis of the ship's motion, determining that there is a significant contribution from rudder motion to rolling motion. This indicated that inappropriate rudder movements caused or at least enhanced rolling motion, but at the same time it also suggested the possibility of controlling and reducing the rolling motion by a proper steering motion. The success of roll-reduction control on ships demonstrates the usefulness of the power contribution analysis in a feedback system. The statistical control method was also applied to the computer control of electric power plants and it was reported that a significant improvement of the control ability was obtained by this method (Nakamura and Akaike (1981)). This controller has been installed in many real electric power plants, not only in Japan, but also in other countries (Nakamura and Akaike (1998)).

The ARMA model became famous, in particular in the econometric area, by the book by Box and Jenkins (1970). However, in that book, a complex method of computing the approximate likelihood was used. In Akaike (1973b), the maximum likelihood method of estimating the parameters of the ARMA model was given. Further, Akaike (1974b) shows a method of determining the ARMA order and initial estimates of the ARMA coefficients based on the Markovian representation of the model. Akaike (1978d) developed a method of computing the variance covariance matrix of the initial state of the ARMA model that enables the computation of the exact likelihood by the Kalman filter. This state space representation became a clue to the systematic use of state space models in time series analysis, such as in the treatment of missing observations (Jones (1980)) and nonstationary time series (Kitagawa and Gersch (1996)).

Prof. Akaike recognized the importance of software and developed a series of time series program packages, TIMSAC (Time Series Analysis and Control). Original TIMSAC was published in 1972 as an appendix of the monograph, Akaike and Nakagawa (1988), and contains the FORTRAN programs for the fitting of, and analysis by the AR model. For identification of the AR model, the Yule-Walker method and FPE criterion are used. TIMSAC-74 contains programs related to the ARMA model and locally stationary AR model (Akaike *et al.* (1975)). TIMSAC-78 contains programs for fitting AR models by the Householder transformation (Akaike *et al.* (1979)). TIMSAC-84 contains programs for Bayesian analysis of nonstationary time series.

5. AIC and statistical modeling

In the 1960's the order selection criterion for autoregressive models, FPE, was proposed based on the predictive point of view. However, to obtain a general model evaluation criterion, we need one more leap. In 1971, Prof. Akaike was aware of the similarity between the MFPE and a criterion for the factor analysis model. However, in factor analysis, it was hard to define the predicted value and the prediction error. He finally realized that, in general, the problem of prediction should be considered by the predictive distribution rather than the predicted values themselves. He also found that the log-likelihood can be considered as an estimate of the Boltzmann's entropy and thus as a measure of the goodness of the predictive distribution (Akaike (1973a), Sakamoto *et al.* (1986), and Konishi and Kitagawa (1996, 2007)). Based on the above considerations, namely, from the following three:

(i) Predictive point of view,
(ii) Prediction by a distribution,
(iii) Evaluation by entropy (or Kullback-Leibler information),

he showed that the log-likelihood is a natural estimate of the expected log-likelihood and, in some sense, that of the K-L information, and therefore can be considered as a measure of goodness of the predictive distribution specified by the assumed model. This clearly reveals the meaning of the maximum likelihood method. He further showed that if we evaluate the model whose parameters are estimated by the maximum likelihood method, then the log-likelihood has a positive bias as an estimator of the expected log-likelihood. By compensating for this bias, he derived the Akaike information criterion (AIC)

$$\begin{aligned} \text{(5.1)} \qquad \text{AIC} &= -2(\text{maximum log-likelihood}) + 2(\text{number of parameters}) \\ &= -2 \log f(x \mid \hat{\theta}) + 2k, \end{aligned}$$

where $\hat{\theta}$ is the maximum likelihood estimate of the parameter vector of the model and k is the number of free parameters or the dimension of the parameter vector. Introduction of the information criterion, AIC, makes it possible to select the order or model in statistical modeling almost automatically, and is used in various fields such as engineering, earth and space sciences, economics, finance, management science, psychology, life science, computer science and information science. Actually the impact of the AIC is very strong and extensively spread (Sakamoto *et al.* (1986), Bozdogan (1994), Akaike and Kitagawa (1998) and Konishi and Kitagawa (2007)), and the year-by-year number of papers citing the two papers advocating the AIC, Akaike (1973a, 1974a), is increasing very rapidly, even after 30 year from their publication. According to the ISI (Thomson) database, the total number of papers citing the above two papers exceed 9000, and in 2006 it was over 10,000, that is an amazing record in the statistical science area.

The AIC criterion had much controversy, in particular, on the consistency of the order of the model. Most of the arguments were based on the misunderstanding of the objective of the modeling posed by Prof. Akaike. According to him, the objective of statistical modeling is to obtain a “good” model for prediction and not to obtain the “true” model. In this context, the consistency is irrelevant, in various senses (Konishi and Kitagawa (2007)). First, the “true” model does not directly relates to the “best” model. Namely, in order selection, for finite samples, the best order may be smaller than the true order, if it exists. Second, if the consistency of the parameters holds, even if the selected order may be larger than the true one, the estimated model becomes consistent. Third, in statistical modeling, the model is built by using various information about the object, past data, and objective of the analysis. In that sense, it is hard to assume the presence of the “true” model and what can assume to exist in the data.

As predicted by Prof. Akaike, various information criteria such as BIC, TIC, c-AIC, RIC, GIC, EIC, are proposed (Konishi and Kitagawa (2007)). It should be noted that, Akaike (1977b) proposed the BIC criterion that is similar to the one proposed later by Schwarz (1978) and is known to have a consistency of order.

6. Bayes modeling

After the proposal of the information criterion, AIC, Prof. Akaike moved to Bayes modeling. In Akaike (1977a, 1978a, b, c, 1979), given the data $Y = \{y_1, \dots, y_n\}$ and the candidate models $M_1, \dots, M_\ell$, he defined the likelihood of the model M_j by $\exp(-\mathrm{AIC}_j/2)$, and the "posterior probability" of the order by

$$p(M_j \mid Y) = \frac{\exp(-\mathrm{AIC}_j/2)}{\sum_{i=1}^{\ell} \exp(-\mathrm{AIC}_i/2)}. \tag{6.1}$$

Then given a properly defined prior probability of the order, π_j, a Bayes estimate of the model is obtained by

$$p(x \mid Y) = \sum_{j=1}^{\ell} p_j(x \mid \hat{\theta}_j) p(M_j \mid Y). \tag{6.2}$$

This type of averaging was done to mitigate the instability intrinsic to the model selection procedure.

However, in the development of the seasonal adjustment method, Akaike (1980a, b) proposed an innovative method of Bayesian modeling. In the standard additive type seasonal adjustment, the observed seasonal time series y_n is decomposed as

$$y_n = T_n + S_n + \varepsilon_n, \tag{6.3}$$

where T_n, S_n and ε_n are the trend, seasonal and irregular components, respectively. In the conventional methods, these components are estimated by either an empirical method or by assuming some parametric models such as a polynomial function for the trend component or a trigonometric function for the seasonal component.

On the other hand, in Akaike (1980b), T_n and S_n are considered parameters of the model. Obviously, it is not possible to obtain meaningful estimates of T_n and S_n by the ordinary least squares method or the maximum likelihood method, since the number of estimated parameters are, at least, $2n$, namely 2 times the number of observations. In the paper, he had to resort to the penalized least squares method that minimizes the criterion

$$\begin{aligned}\sum_{i=1}^{n} [(y_i - T_i - S_i)^2 &+ d^2(T_i - 2T_{i-1} + T_{i-2})^2 \\ &+ r^2(S_i - S_{i-12})^2 + z^2(S_i + \cdots + S_{i-11})^2],\end{aligned} \tag{6.4}$$

where, d, r and z are properly chosen constants. This type of penalized least squares method, or the penalized maximum likelihood method, have been used in Whittaker (1923) and Good and Gaskins (1980). However, the selection of the crucial parameters d, r and z are left to the discretion of the analysts. Akaike (1979, 1980a) showed a Bayesian interpretation of the above criterion. Namely, by putting $y = (y_1, \ldots, y_n)^T$, $a = (T_1, \ldots, T_n, S_1, \ldots, S_n)^T$, the above criterion can be expressed as

$$\|y - Ha\|^2 + \lambda^2 \|Da\|^2, \tag{6.5}$$

where $\| \cdot \|$ is a properly defined norm of a vector. Then by multiplying $-1/2\sigma^2$ and taking the exponential, we obtain

$$\exp\left\{-\frac{\|y - Ha\|^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{\lambda^2\|Da\|^2}{2\sigma^2}\right\}. \tag{6.6}$$

Since these two terms can be considered the principal part of the normal densities, they can be interpreted as the data distribution and the prior distribution of a Bayesian model, respectively. Based on this interpretation, Bayesian information criterion (ABIC) was defined by (Akaike (1980a))

$$\begin{aligned} \text{ABIC} = &-2(\text{maximum log likelihood of a Bayesian model}) \\ &+ 2(\text{number of estimated hyper-parameters}). \end{aligned} \tag{6.7}$$

This provides us with a practical method of determining the prior distribution of the Bayes model. There have been two difficulties in developing practical Bayes procedures, namely the selection of the prior distribution and the computation of the posterior distribution. Due to the development of fast computers and various computational methods such as MCMC and sequential Monte Carlo methods, Bayes modeling became practical and the key technology in various fields of scientific research and intellectual information processing. As important applications of this Bayes modeling, we may consider the seasonal adjustment method (Akaike and Ishiguro (1980), Kitagawa and Gersch (1984)), time-varying spectrum estimation (Kitagawa (1983)), tidal data analysis (Ishiguro *et al.* (1983)), factor analysis (Akaike (1987)), cohort analysis (Nakamura (1986)), seismic data analysis (Kitagawa and Takanami (1985), Kitagawa and Matsumoto (1996)), data assimilation (Fukuda *et al.* (2004)), etc. Some other applications can be seen in Akaike and Kitagawa (1998) and Higuchi (2007).

After retirement from the Institute of Statistical Mathematics in 1994, where he acted as the Director-General for 8 years, he concentrated on the analysis of golf swing motion (Akaike (2001, 2003)) as an example of statistical thinking. In Akaike (2001), he claimed that there are three types of informational data sets composed of objective knowledge, empirical findings, and observational data, and that we should fully utilize these in building statistical models.

Acknowledgements

The author is grateful to the referee for his valuable comments.

References

Akaike, H. (1956). On a zero-one process and some of its applications, *Ann. Inst. Statist. Math.*, **8**, 87–94.

Akaike, H. (1959a). On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method, *Ann. Inst. Statist. Math.*, **11**, 1–16.

Akaike, H. (1959b). On the statistical control of the gap process, *Ann. Inst. Statist. Math.*, **10**, 233–259.

Akaike, H. (1962). On the design of lag window for the estimation of spectra, *Ann. Inst. Statist. Math.*, **14**, 1–21.

Akaike, H. (1964). Statistical measurement of frequency response function, *Supplement III*, *Ann. Inst. Statist. Math.*, 5–17.

Akaike, H. (1967). Some problems in the application of the cross-spectral method, *Spectral Analysis of Time Series* (ed. B. Harris), 81–107, John-Wiley.

Akaike, H. (1968). On the use of a linear model for the identification of feedback systems, *Ann. Inst. Statist. Math.*, **20**, 425–439.

Akaike, H. (1969a). Fitting autoregressive model for prediction, *Ann. Inst. Statist. Math.*, **2l**, 243–247.

Akaike, H. (1969b). Power spectrum estimation through autoregressive model fitting, *Ann. Inst. Statist. Math.*, **21**, 407–419.

Akaike, H. (1970). Statistical predictor identification, *Ann. Inst. Statist. Math.*, **22**, 203–217.

Akaike, H. (1971). Autoregressive model fitting for control, *Ann. Inst. Statist. Math.*, **23**, 163–180.

Akaike, H. (1973a). Information theory and an extension of the maximum likelihood principle, *Proc. 2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), Akademiai Kiado, Budapest, 267–281. (Reproduced in Breakthroughs in Statistics, **I**, Foundations and Basic Theory, S. Kotz and N. L. Johnson eds., Springer-Verlag, New York, (1992), 610–624.)

Akaike, H. (1973b). Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, **60**, 255–265.

Akaike, H. (1974a). A new look at the statistical model identification, *IEEE Trans. Automatic Control*, **AC-19**(6), 716–723.

Akaike, H. (1974b). Stochastic theory of minimal realization, *IEEE Trans. Automatic Control*, **AC-19**(6), 667–674.

Akaike, H. (1977a). An objective use of Bayesian models, *Ann. Inst. Statist. Math.*, **29**, 9–20.

Akaike, H. (1977b). On entropy maximization principle, *Applications of Statistics* (ed. P. R. Krishnaiah), 27–41, North-Holland Publishing Company.

Akaike, H. (1978a). A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, **30**A, 9–14.

Akaike, H. (1978b). A new look at the Bayes procedure, *Biometrika*, **65**, 53–59.

Akaike, H. (1978c). On the likelihood of a time series model, *The Statistician*, **27**, 217–235.

Akaike, H. (1978d). Covariance matrix computation of the state variable of a stationay Gaussian process, *Ann. Inst. Statist. Math.*, **30**B, 499–504.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, **66**(2), 237–242.

Akaike, H. (1980a). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith), 143–166, University Press, Valencia, Spain (discussion 185–203).

Akaike, H. (1980b). Seasonal adjustment by a Bayesian modeling, *Journal of Time Series Analysis*, **1**(1), 1–13.

Akaike, H. (1987). Factor analysis and AIC, *Psychometrika*, **3**, 317–332.

Akaike, H. (2001). Golf swing motion analysis: an example of verbal analysis in statistical reasoning, *Ann. Inst. Statist. Math.*, **53**, 1–10.

Akaike, H. (2003). On the art of modeling; Illustrated with the analysis of the golf swing motion, *Genome Informatics*, **14**, 263–265.

Akaike, H. and Ishiguro, M. (1980). BAYSEA, A Bayesian Seasonal Adjustment Program, *Computer Science Monographs*, **13**, The Institute of Statistical Mathematics, Tokyo.

Akaike, H. and Kitagawa, G. (1998). *The Practice of Time Series Analysis*, Springer-Verlag, New York.

Akaike, H. and Nakagawa, T. (1988). *Statistical Analysis and Control of Dynamic Systems*, Kluwer Academic Pub., Dordrecht.

Akaike, H. and Yamanouchi, Y. (1962). On the statistical estimation of frequency response function, *Ann. Inst. Statist. Math.*, **14**, 23–56.

Akaike, H., Arahata, E. and Ozaki, T. (1975). *TIMSAC-74*, *Computer Science Monographs*, No. 5, The Institute of Statistical Mathematics, Tokyo.

Akaike, H., Kitagawa, G., Arahata, E. and Tada, F. (1979). *TIMSAC-78, Computer Science Monographs*, No. 11, The Institute of Statistical Mathematics, Tokyo.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis*, *Forecasting and Control*, Holden-Day, San Francisco.

Bozdogan, H. (ed.) (1994). *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling*: *An Informational Approach*, Kluwer Academic Publishers, the Netherlands.

Burg, J. P. (1968). Maximum entropy spectral analysis, *Modern Spectrum Analysis*, 34–48, IEEE Press.

Fukuda, J., Higuchi, T., Miyazaki, S. and Kato, T. (2004). A new approach to time-dependent inversion of geodetic data using Monte Carlo mixture Kalman filter, *Geophysical Journal International*, **159**, 17–39.

Good, I. J. and Gaskins, R. A. (1980). Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data, *Journal of the American Statistical Association*, **75**, 42–56.

Higuchi, T. (ed.) (2007). *Statistical Mathematics Reveals Hidden Future*: *Innovation in the Real World by Bayesian Modeling*, Tokyo Denki University, Press, Tokyo (in Japanese).

Ishiguro, M., Ooe, H., Nakai, S. and Akaike, H. (1983). A Bayesian approach to the analysis of earth tides, *Proceedings of the Ninth International Symposium on Earth Tides* (ed. J. T. Kuo), 283–292.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, **22**, 389–395.

Kitagawa, G. (1983). Changing spectrum estimation, *Journal of Sound and Vibration*, **89**(4), 433–445.

Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *Journal of the American Statistical Association*, **79**, 378–389.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, **116**, Springer-Verlag, New York.

Kitagawa, G. and Matsumoto, N. (1996). Detection of coseismic changes of underground water level, *Journal of the American Statistical Association*, **91**, 521–528.

Kitagawa, G. and Takanami, T. (1985). Extraction of a signal by a time series model and screening out micro earthquakes, *Signal Processing*, **8**, 303–314.

Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection, *Biometrika*, **83**(4), 875–890.

Konishi, S. and Kitagawa, G. (2007). *Information Criteria and Statistical Modeling*, Springer-Verlag, New York.

Kowarik, J. and Osborne, M. R. (1968). *Methods for Unconstrained Optimization Problems*, Elsevier, New York.

Nakamura, H. and Akaike, H. (1981). Statistical identification for optimal control of supercritical thermal power plants, *Automatica*, **17**(1), 143–155.

Nakamura, H. and Akaike, H. (1998). Control of boilers for thermoelectric power plants by means of a statistical model, *The Practice of Time Series Analysis* (eds. H. Akaike and G. Kitagawa), 1–17, Springer-Verlag, New York.

Nakamura, T. (1986). Bayesian cohort models for general cohort table analysis, *Ann. Inst. Statist. Math.*, **38**(2), 353–370.

Otomo, T., Nakagawa, T. and Akaike, H. (1972). Statistical approach to computer control of cement rotary kilns, *Automatica*, **8** 35–48.

Otsu, K., Horigome, M. and Kitagawa, G. (1979). A new ship's auto pilot design through a stochastic model, *Automatica*, **15**(3), 255–268.

Parzen, E., Tanabe, K. and Kitagawa, G. (eds.) (1998). *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics, Springer-Verlag, New York.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*, D. Reidel, Dordrecht.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

Whittaker, E. (1923). On a new method of graduation, *Proceedings of Edinburgh Mathematical Society*, **41**, 63–75.