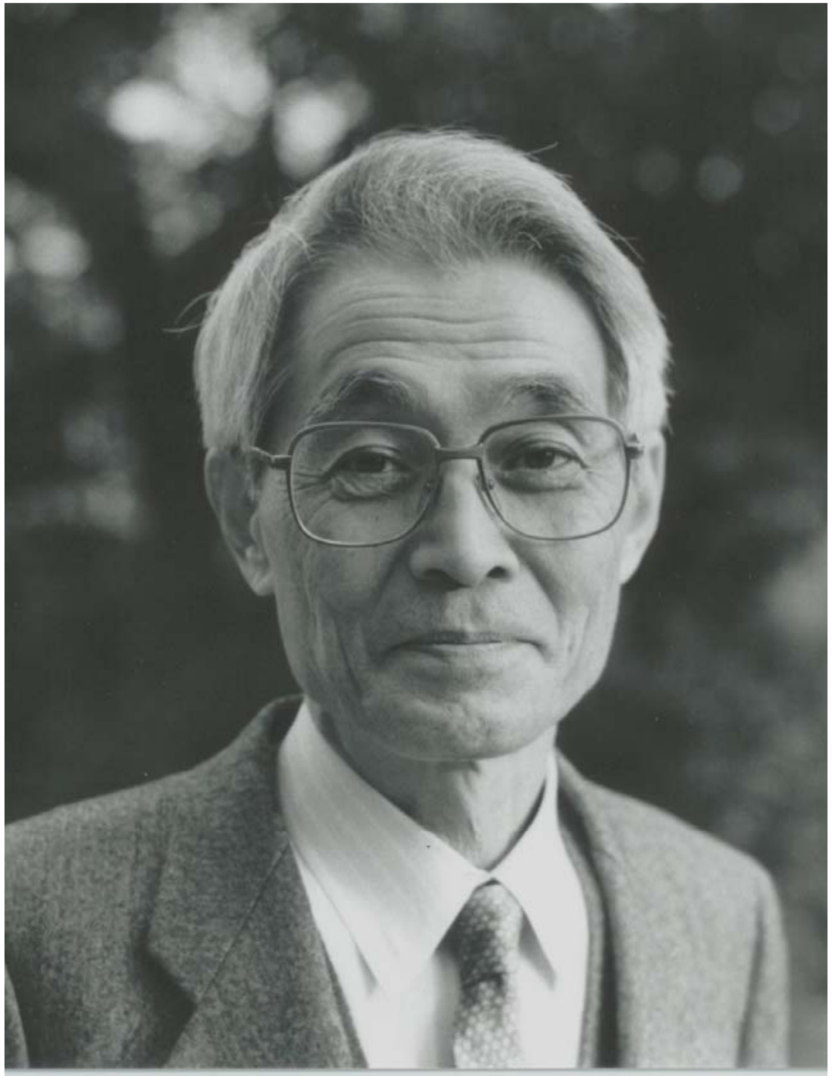
Contribution of Prof. H. Akaike to Statistical Modeling of Complex Systems

Research Organization of Information and Systems

Genshiro Kitagawa

Hirotugu Akaike（1927－2009）



赤池弘次

Brief History

1952 Graduated from Math. Dept., Tokyo University
Researcher of the **Institute of Statistical Mathematics**

1962 Head of 2nd Section, 1st Division

1973 Director of **5th Division**

1985 Director of **Dept. of Prediction and Control**

1986 **Director-General of ISM** (-1994)

1988 Member of **Science Council of Japan** (- 1991)
Head of Dept. of Statistical Science, Graduate University of Advanced Study

1994 Prof. Emeritus, The Inst. Statist. Math.
Prof. Emeritus, Graduate Univ. for Advance Study

Prizes

1972 Ishikawa Prize

(Establishment of statistical analysis and control method for dynamic systems)

1980 Okochi Prize

(Research and realization of optimal steam temperature control of thermal electric plant)

1989 Asahi Prize

(Research on statistics, in particular theory and applications of AIC)

The Purple Ribbon Medal

(Statistics, in particular time series analysis and its applications)

1996 The 1st Japan Statistical Society Prize

(Contributions to statistical theory and its applications)

2000 The Order of the Sacred Treasure

2006 Kyoto Prize

(Major contribution to statistical science and modeling with the development of AIC)

Fellow of ASA, RSS, IMS, IEEE, JSS

Laureate of 22nd **Kyoto Prize**

稲盛財団
INAMORI FOUNDATION

2006.6.9

プレスリリース

第 22 回（2006）京都賞受賞者の決定

財団法人稲盛財団（理事長・稲盛和夫）は第 22 回（2006）京都賞の受賞者を決定しました。
本年の受賞者は、以下の 3 名です。

■先端技術部門
本年授賞対象分野：バイオテクノロジー及びメディカルテクノロジー
レナード・アーサー・ハーツェンバーグ博士（アメリカ・74 歳・スタンフォード大学教授）

■基礎科学部門
本年授賞対象分野：数理科学
赤池弘次博士（日本・78 歳・統計数理研究所名誉教授）

"Major contribution to statistical science and modeling with the development of the Akaike Information Criterion (AIC)”

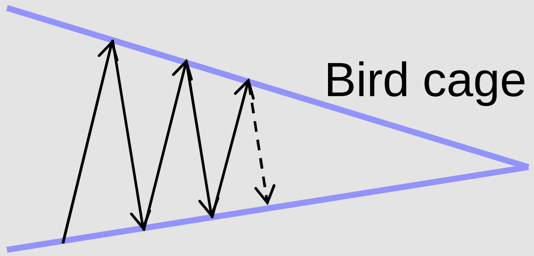
Ontline

1. Launching period: Mathematical analysis and structural modeling (1950's)
2. Frequency domain analysis (1960's)
3. Time domain time series modeling (1970's)
4. AIC and statistical modeling (1970's)
5. Bayes modeling (1980's)

1. Launching Period (1952-1960)

Mathematical Analysis

- Decision process
- Evaluation of probability distribution
- Monte Carlo method solving linear equation
- Computation of eigenvalues
- Convergence of optimum gradient method



Convergence Analysis of Optimal Gradient Method **AISM**(1959)
Analyzed limiting behavior of the optimal gradient method and showed the poor convergence property, known as **bird cage**.

Became the **foundation of nonlinear optimization methods** such as Conjugate Gradient method and Quasi Newton method

… Kowarik and Osbone (1968)

Structural Modeling

At this time, he was rather interested in real-world problems. And from very early stage in the1950', he realized the conventional linear stationary models are unrealistic and the importance of developing a model that fully takes into account of the structure of the process.

Control of filature production process

(Joint work with Dr. Shimazaki of the Sericultural Experiment Station)

- Developed a control method based on gap process modeling
- This method provided a reference process to detect abnormalities in actual reeling process.
- Brought significant innovation for silk production in Japan

2. Frequency Domain Analysis: (1960-70)

- By 1960, he established contacts with engineers

> It took almost 10 years for me to develop substantial contacts, and this happened largely by coincidence. Statistical Science (1995)

- He realized that linear stationary modeling is utilized and many unsolved important problems are left for statisticians
- Smoothing method for power spectrum estimation
  (Suspension system of a car, Dr. Kaneshige, Isuzu Motor Co.)
  Akaike windows
- Estimation of frequency response function
  (Dr. Yamanouchi, Transportation Tech. Res. Inst.)
  Developed a method of estimating frequency response function from observations under normal steering (without using sinusoidal inputs)

2. Frequency Domain Analysis: (1960-70)

Developed practical method of analyzing complex systems

Organized workshop on practical use of time series method

- Ship's yawing, rolling (Yamanouchi, Kawashima)
- Rolling of a car （Kawamura)
- Engine of a car （Kaneshige)
- Response of an air-plane to side wind （Takeda)
- Hydroelectric power plant （Nakamura）
- Estimation of underground structure through micro tremor
- Tsunami (Kinosita）
- EEG analysis (Suhara)

AISM supplement (1964)

Difficulty in Feedback Systems

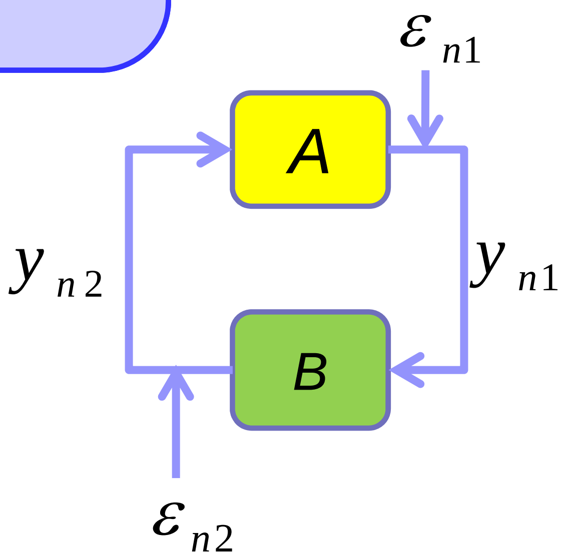
Cement Rotary Kiln, Chichibu Cement Company

- rotary kiln is a complicated **feedback system**, consisting of many variables such as raw material feed, fuel feed, gas damper angle, temperature at various locations, etc.
- Conventional spectral analysis methods were useless to identify the source of fluctuation
- In the frequency domain analysis, we cannot explicitly utilize the physical realizability of the system

Limitation of frequency domain approach

Madison symposium (1967)

→ Return to time domain modeling



3. Time Domain Time Series Modeling (1968-)

Vector time series $y_n = (y_{n1}, \ldots, y_{n\ell})^T$

Structural model

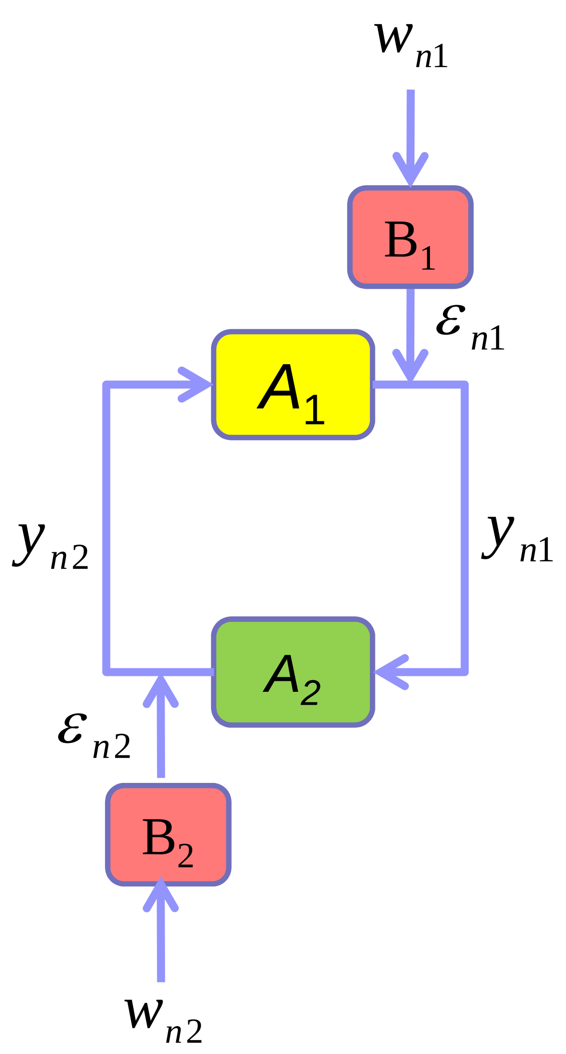
$$y_{ni} = \sum_{j \neq i} \sum_{k=1}^{m} \alpha_{kij} y_{n-k,j} + \varepsilon_{ni}$$

$$\varepsilon_{ni} = \sum_{k=1}^{\ell} \beta_{ki} \varepsilon_{n-k,i} + w_{ni}$$

VAR model representation

$$y_n = \sum_{k=1}^{m+\ell} A_k y_{n-k} + w_n$$

Estimation of α_{kij} and β_{ki} through A_k



Spectrum & Power Contribution

VAR model

$$y_n = \sum_{k=1}^{m} A_k y_{n-k} + w_n \qquad w_n \sim N(0,\Sigma)$$

Cross spectrum

$$p(f) = B(f)\,\Sigma\,B(f)^*$$

$$B(f) = (b_{ij}) = \sum_{\ell=1}^{m} A_\ell \exp(-2\pi i \ell f)$$

Power spectrum of y_{ni}

$$p_{ii}(f) = \sum_{j=1}^{\ell} |b_{ij}(f)|^2 \sigma_j^2$$

Power contribution

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

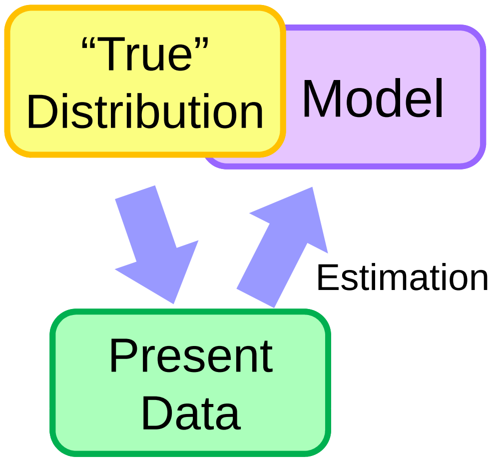
Order m is unknown!

Proper selection of the order is crucial

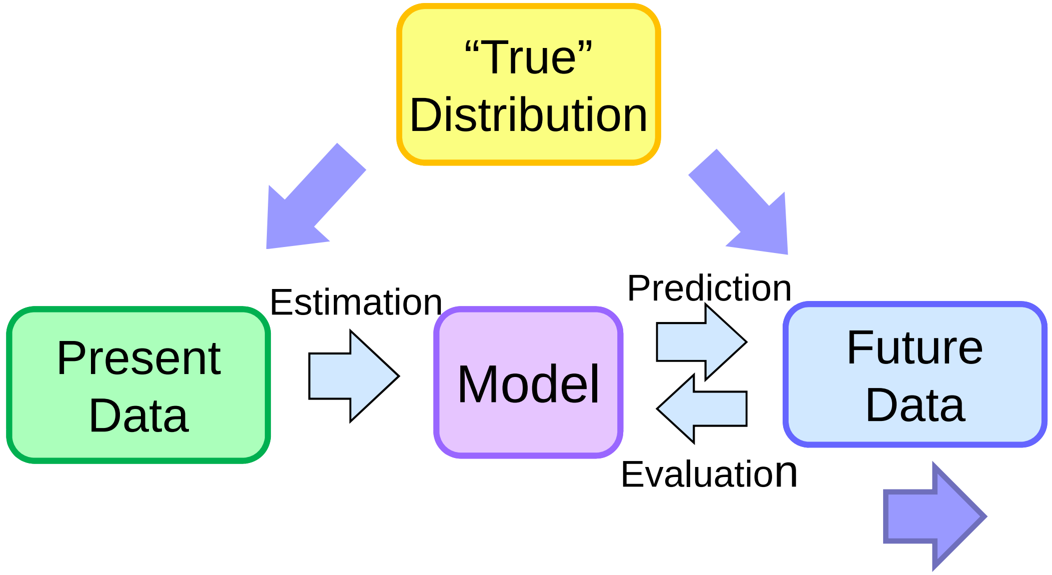
Proportion of the contribution from noise input $e_j(n)$ in the power spectrum of $y_i(n)$ at frequency f.

Predictive Point of View and FPE

Conventional fitting view point



Predictive view point



Scalar AR model

$$y_n = \sum_{k=1}^{m} a_k y_{n-k} + w_n, \qquad w_n \sim N(0, \sigma_m^2)$$

FPE (Final Prediction Error)

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

Multivariate case: MFPE, FPEC

FPE is a forerunner of AIC

$$n \log \mathrm{FPE}_m = n \log \hat{\sigma}_m^2 + n \log\left(\frac{n+m+1}{n-m-1}\right)$$

$$\cong n \log \hat{\sigma}_m^2 + (m+1)$$

Realization of Statistical Control

VARX model

$$y_n = \sum_{j=1}^{m} A_j y_{n-j} + \sum_{j=1}^{m} B_j r_{n-j} + \varepsilon_n$$

State –Space Representation

$$x_n = F x_{n-1} + G r_n + \varepsilon_n$$

$$y_n = H x_n$$

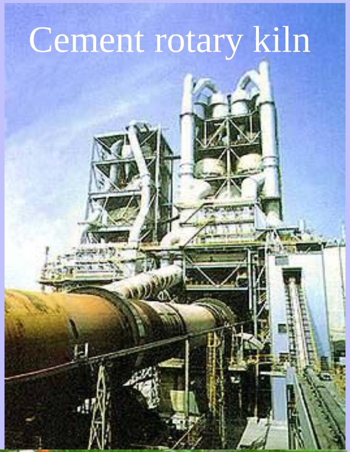
Criterion Function

$$J = \sum_{j=1}^{L} E\left(x_n^T Q x_n + r_n^T R r_n\right)$$

Statistical Optimal controller

$$r_n^* = K x_n + H r_{n-1}$$

TIMSAC Akaike-Nakagawa(1972)



Nakamura-Akaike(1981)

Ohtsu-Horigome-Kitagawa(1978)

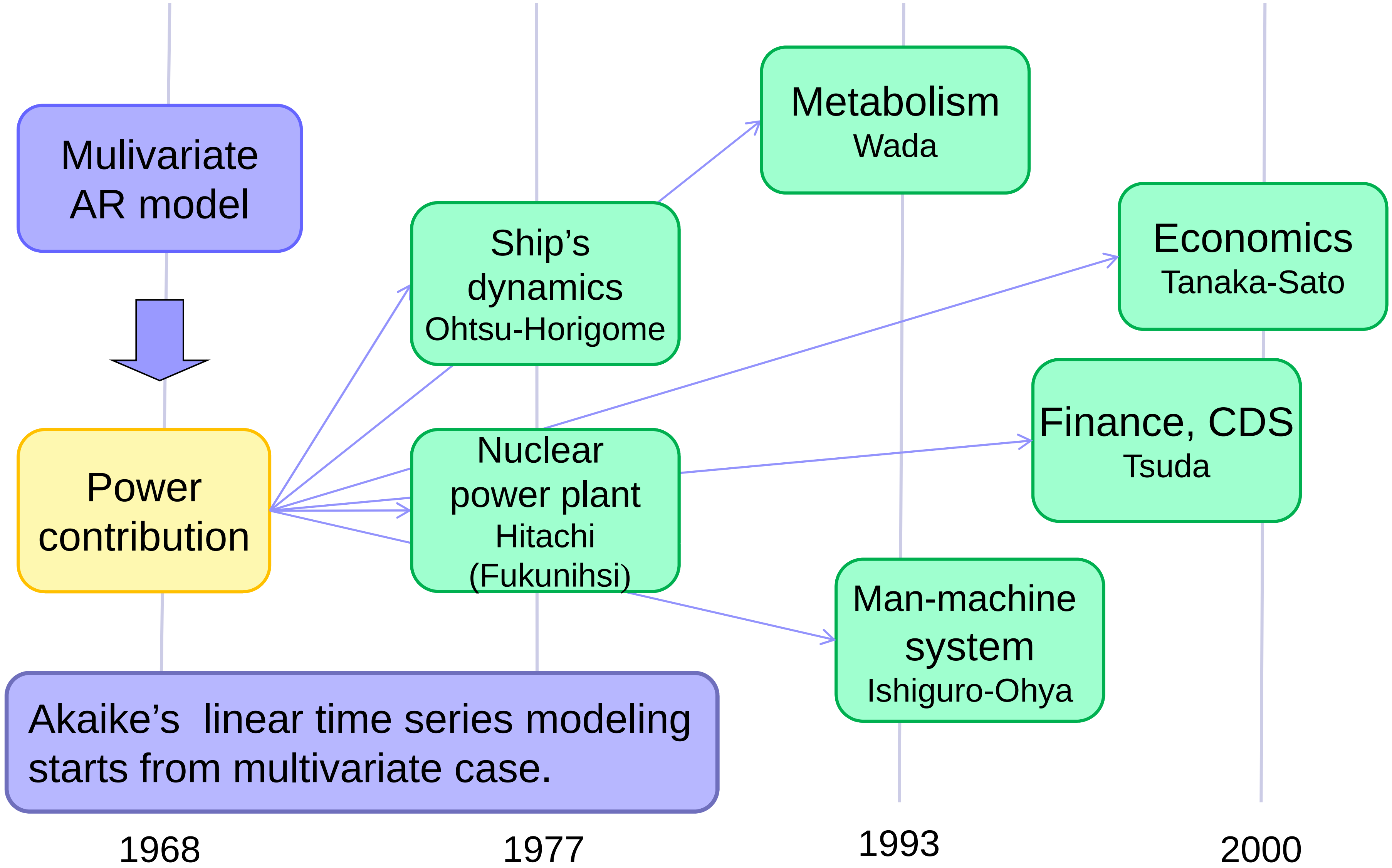
Nonlinear control
Japan Bailay
Ozaki-Peng

Roll-reducing control
Mitsui ship building
Oda

Noise-adaptive controller
Yokokawa
Kitagawa-Akaike-Ohtsu (2006)

1970 1980 2000

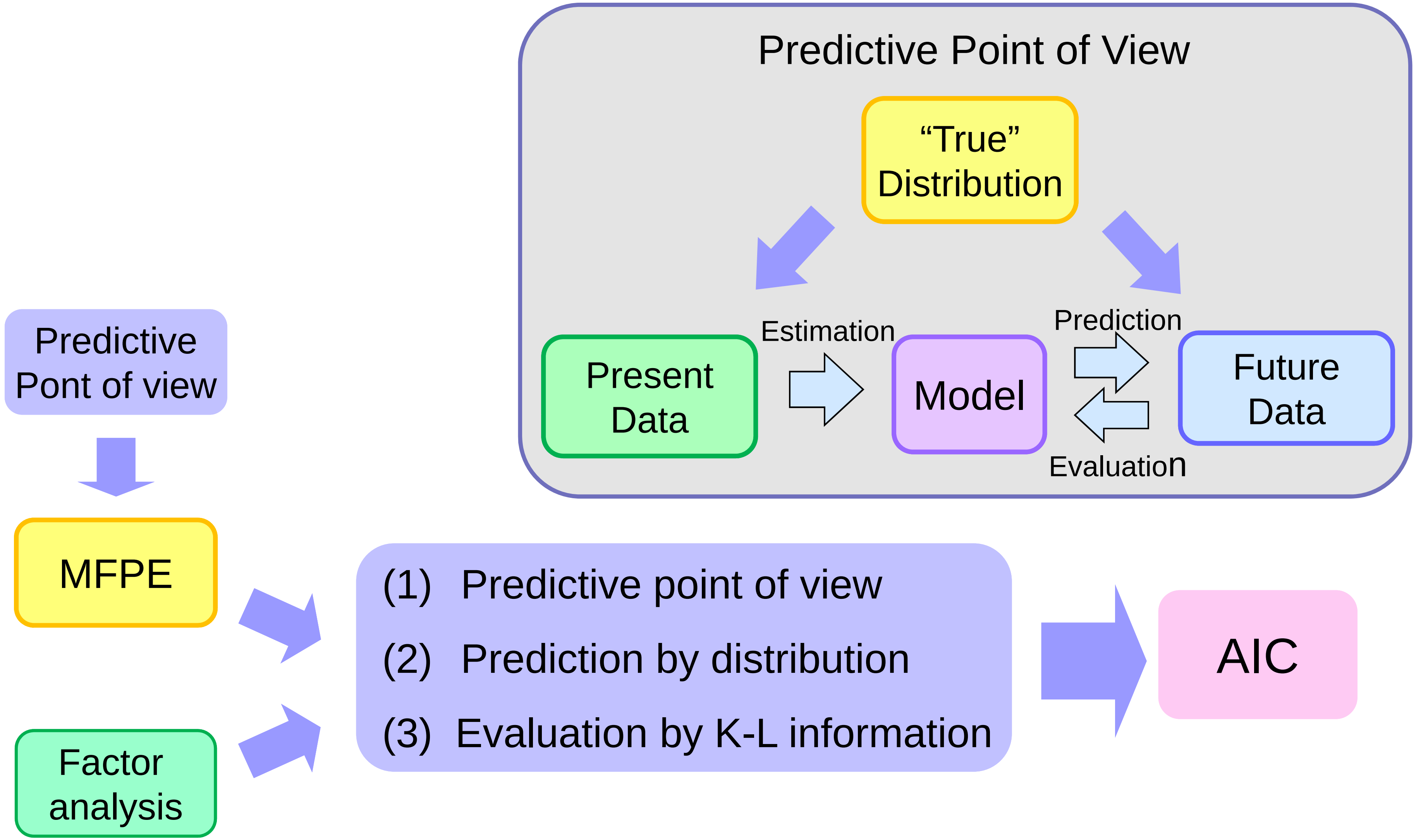
Analysis and Control of Feedback Systems



Main Contributions in Time Domain Modeling

1. Developed practical method of VAR modeling
   - Analysis of feedback systems
   - Statistical controller
2. Use of state-space model
   - For optimal controller design
   - ML Estimation and identification of VARMA model
   - BAYSEA: Bayesian method for seasonal adjustment
3. Oder determination
   - FPE, MFPE, FPEC
   - AIC
4. Developed TIMSAC (Time Series Analysis and Control program package)

4. AIC and Statistical Modeling (1971-)



Predictive Point of View and AIC

Model $g(y)$: distribution of future data, $f(y)$: model

Kullback-Leibler Information

$$I(g;f) = E_Y \log\left(\frac{g}{f}\right) = E_Y \log g - E_Y \log f$$

Unknown constant

Expected Log-Likelihood

$$E_Y \log f(Y) = \int \log f(y) g(y) dy$$

Log-Likelihood $x_1, \ldots, x_n \sim g(x)$

$$\frac{1}{n}\sum_{i=1}^{n} \log f(X_i) \longrightarrow E_Y \log f(Y)$$

MLE

$$\ell(\theta) \equiv \log f(X \mid \theta) \qquad \max \ell(\theta) \Rightarrow \hat{\theta} = \hat{\theta}(X)$$

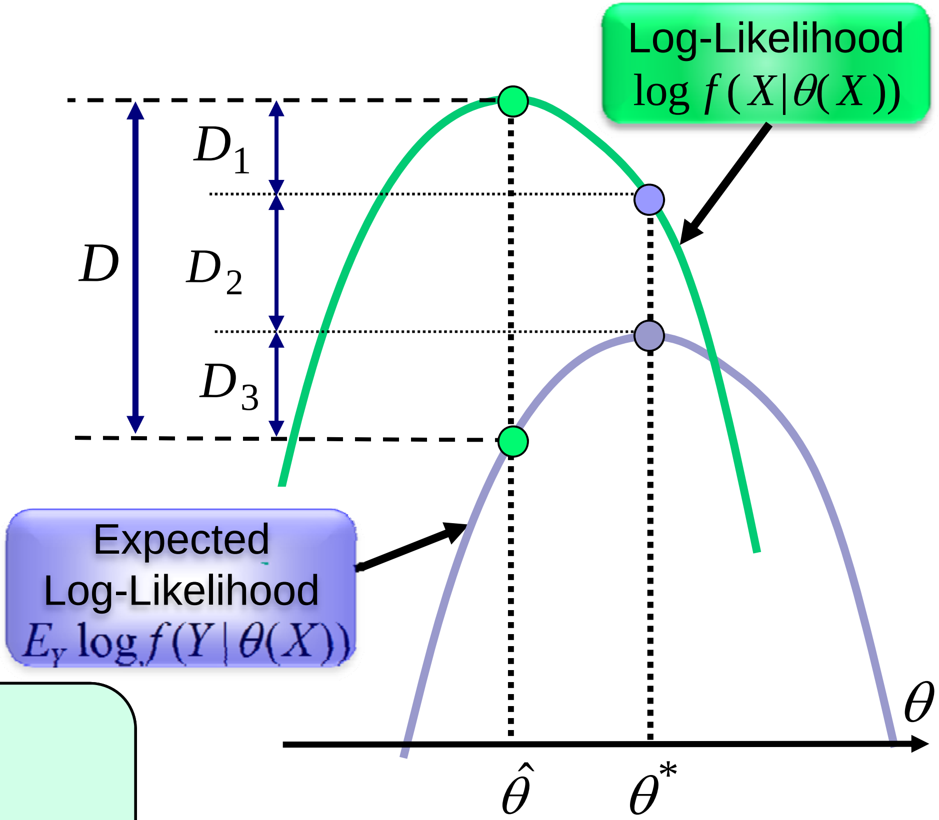
Multi-model Situation: Bias correction ⇒ **AIC**

Bias Correction

$$b(G) = E_X\left\{\log f(X \mid \hat{\theta}(X)) - nE_Y \log f(Y \mid \hat{\theta}(X))\right\} = E_X[D]$$

Generic Information Criterion

$$\log\ f(X \mid \hat{\theta}(X)) - b(G)$$



$$b(G) = \mathrm{tr}\left\{I(G)J(G)^{-1}\right\} \cong m$$

$I(G)$: Fisher Information
$J(G)$: -(Expected Hessian)

$$\mathrm{AIC} = -2\log f(X \mid \hat{\theta}(X)) + 2m$$
$$= -2(\log L(\hat{\theta})) + 2(\dim \text{ of } \theta)$$

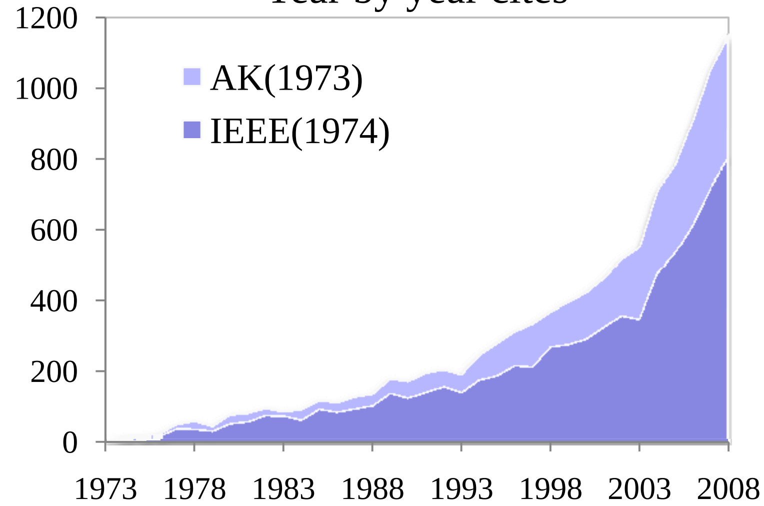
Number of papers citing Akaike's paper by Thomson (ISI)

5142	IEEEac(1974)
2907	Akad. Kiado (1973)
1115	Psychometrika (1987)
818	AISM(1969)
431	AISM(1970)
359	Math. Science (1978)
281	Appl. Stat. (1977)
265	Bayesian Stat. (1980)
204	System Ident. (1980)
135	Celebration (1986)
127	AISM(1969)
123	AISM(1974)
13104	total

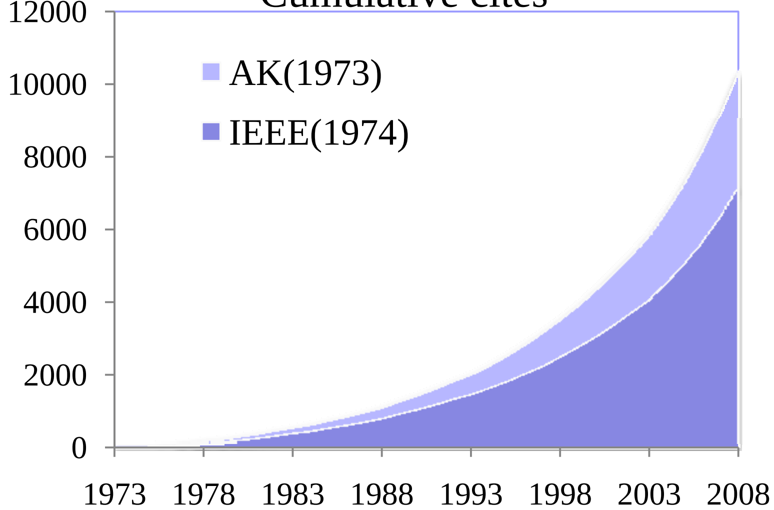
By **Google Scholar**

14265	IEEE-ac (1974)
8588	Acad. Kiado (1973)
1987	Psychometrika(1987)
1969	AISM (1969)

Year by year cites



Cumulative cites



Cited Research Area

2550	Math. Stat.
1987	Medical, Health
1868	Biology
1833	Engineering
1755	Computer, Info. Sci.
1376	Psychology
1070	Economics
785	Social Science
773	Environmental Sci.
673	Physics, Chemistry
610	Geophysics
561	Pharmacology
1428	Others

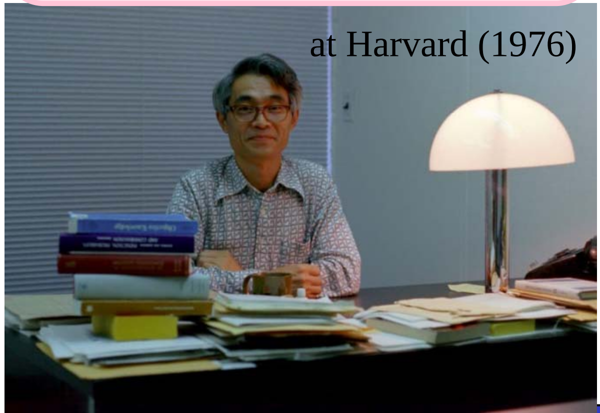
Main Contributions in This Stage

1. Developed of AIC
2. Importance of statistical modeling
3. Motivated nonstationary, non-Gaussian nonlinear modeling
4. Developed TIMSAC-78

5. From AIC to Bayes Modeling

AIC → MAICE

Prof. Akaike shifted to Bayesian modeling by 1976.



Models: $M_1, \ldots, M_\ell$ $\quad \mathrm{AIC}_1, \ldots, \mathrm{AIC}_\ell$

Likelihood of the model: $\exp\left(-\mathrm{AIC}_j / 2\right)$

Prior probability of order: π_j

Posterior probability of order:

$$p(M_j \mid Y) = \frac{\pi_j \exp\left(-\mathrm{AIC}_j / 2\right)}{\sum_{j=1}^{\ell} \pi_j \exp\left(-\mathrm{AIC}_j / 2\right)}$$

Bayes estimate of the model

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

Model averaging mitigated the instability inherent in model selection

Smoothness Prior : Bayesian Seasonal Adjustment

Seasonal Adjustment Problem

$$y_n = T_n + S_n + \varepsilon_n, \quad n = 1, \ldots, N$$

y_n	Observation	S_n	Seasonal component
T_n	Trend	ε_n	Irregular component

Penalized Least Squares

$$\min_f \sum_{n=1}^{N} \Big[(y_n - T_n - S_n)^2 + d^2 (\nabla^2 T_n)^2 + r^2 (S_n - S_{n-12})^2 + z^2 (S_n + \cdots + S_{n-11})^2 \Big]$$

Infidelity to the data

Infidelity to smoothness

Whittaker (1923), Shiller (1973), Good-Gaskins (1980), Akaike(1980),

Determination of Trade-off Parameter via Bayesian Interpretation

(d^2, r^2, z^2) — Crucial parameters

$$\sum_{n=1}^{N}\left[(y_n - T_n - S_n)^2 + \sum_{n=1}^{N}\left[d^2(\nabla^2 T_n)^2 + r^2(S_n - S_{n-12})^2 + z^2(S_n + \cdots + S_{n-11})^2\right]\right.$$

Multiply by $-1/(2\sigma^2)$ **and exponentiate**

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - T_n - S_n)^2\right\}\exp\left\{-\frac{d^2}{2\sigma^2}\sum_{n=1}^{N}(\nabla^2 T_n)^2 + \cdots\right\}$$

Bayesian Interpretation $\theta = (d^2, r^2, z^2, \sigma^2)$

$$\pi(T,S \mid y,\theta) \propto p(y \mid T,S,\theta)\,\pi(T,S \mid \theta)$$

Smoothness Prior (for $\pi(T,S \mid \theta)$)

$$\text{ABIC} = -2(\max \log L) + 2(\dim \text{ of } \theta)$$

Determination of θ by ABIC (Akaike 1980)

Various Practice of Bayes Modeling

Seasonal Adjustment
- BAYSEA
- DECOMP

Earth-Tide Analysis
- BAYTAP-G

Cohort Analysis

State-Space Modeling
- Time-varying AR model
- Extraction of effect of earthquake from groundwater level data
- Nonlinear smoothing
- Data assimilation

Software

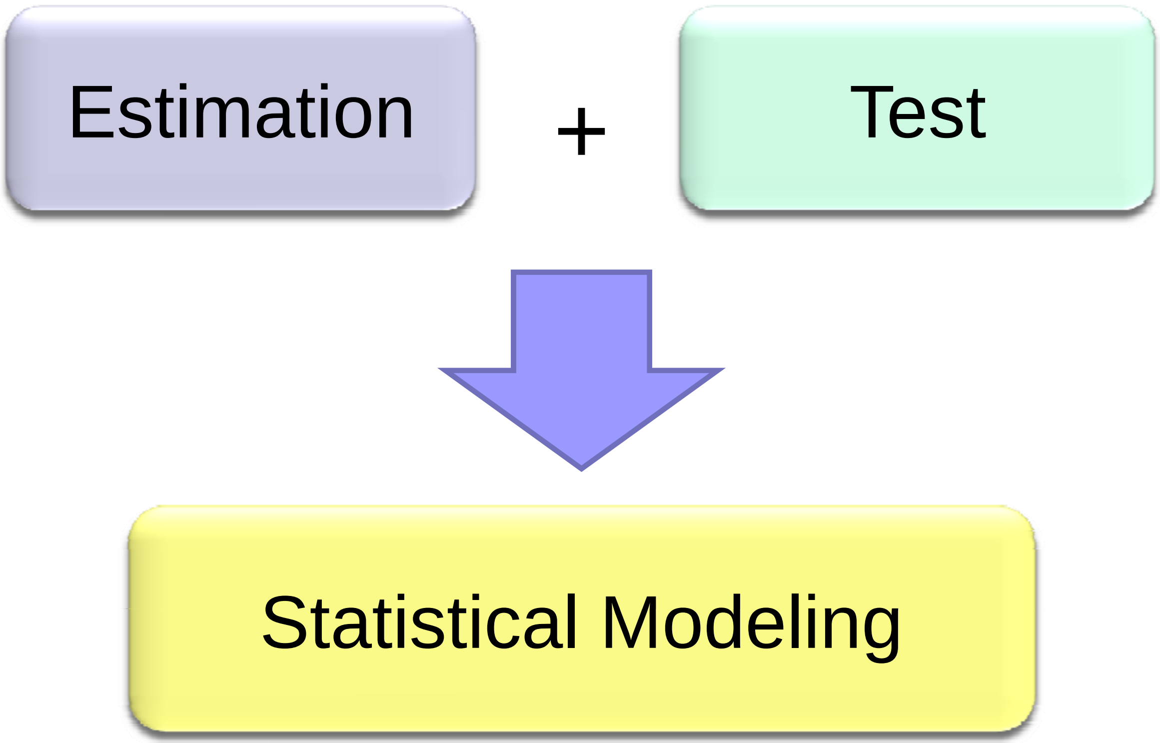
TIMSAC84

Time series analysis programs based on Bayesian modeling

Akaike and Kitagawa (1998), Higuchi (2007).

Impacts of AIC and ABIC

Shift of Statistical Paradigm



Data in Information & Knowledge Society

Information Society
Information Technology

Information Technologies
- **Measurement devices**
- **Internet communication**
- **Database**

Accumulation of Data

- **Life Science**: DNA data, Micro-array data
- **Marketing**: POS data
- **Finance**: High frequency data
- **Environmental Science**:
- **Seismology, Meteorology**
- **Astronomy** (Whole-sky CCD camera)
- **High energy physics**

Massive Data

Statistical Modeling

Data-centric Science （Fourth Science）

	Human Inspiration dependent	**Cyber-enabled**
Deductive （Principle driven）	Theoretical science	Computing Science （Simulation）
Inductive （Data driven）	Experimental Science	Data-centric Science (Fourth science）

Main Contributions of Prof. Akaike

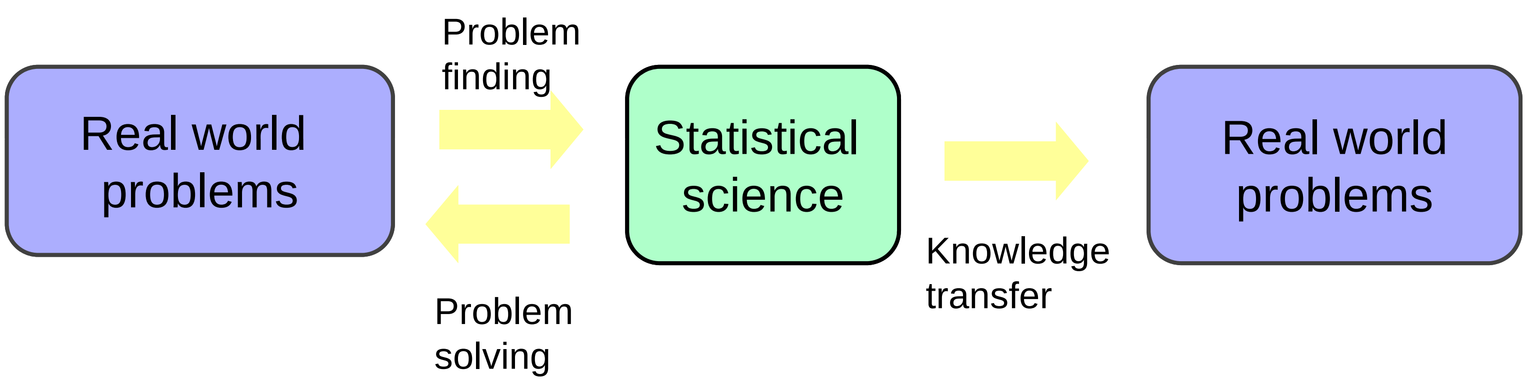
1. Time domain multivariate modeling
   Identification method and various practices

2. Information criteria
   Change of statistical paradigm: modeling

1984

Concluding Remarks: Akaike's Research Style

- Challenge real problems for problem finding
- Joint works with researchers in various domains
- Occasionally change his policy if necessary
- Develop and open software



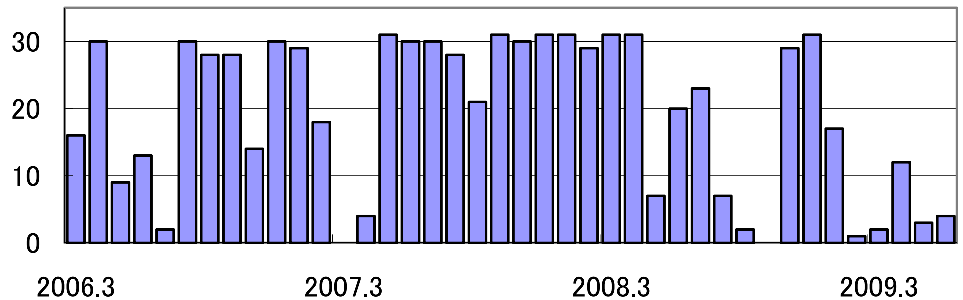
Epilogue: Analysis of Golf Swing (1994 - 2009)

Golf swing

Golf swing motion analysis: An experiment on the use of verbal analysis in statistical reasoning, **AISM**, Vol.53, No.1 1-10 (2001).

Blog http://ameblo.jp/linear/

Renewal of his Blog



Plausibility

Evaluation and comparison of verbally expressed psychological or physical image

Probability > Likelihood > Plausibility

→ Statistical Thinking

Making statistical thinking more productive, to appear in **AISM**, Vol.62, No.1 (2010).