From Data Analysis to Data Science

An Overview and Future Prospects: the Research Interchange in Data Analysis Between Japan and France

SFC-97, Lyon
France

Noboru Ohsumi

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

In my talk today, ...

(1) To present a brief overview of the research interchange in the field of *Data Analysis* between Japan and France.

(2) To introduce you the history of the interaction between France and Japan from my perspective over quite a long period.

It must be remembered that in both Japan and France, further fruitful fields of data analysis have been cultivated, separately from the development of EDA proposed by J. Tukey.

It has been nearly 20 years since the bonds of relationship were formed between Japan and France.

We cannot help being surprised at the fact that ideas which turned out to be very similar to each other were born in such distant and different cultures almost at the same time.

It is a pleasure to have this opportunity to acquaint you with the past, present, and future of this important and productive research exchange in data analysis between the two cultures.

1. The Beginning of the Research Interchange between Japan and France

• The research interchange in Japan and France has a longer history.

• However, this history is not widely known among the researchers here in France.

Japanese-French Scientific Seminar was held at The University of Paris VI (Université Pierre-et-Marie Curie) , 1978.

Theme:

Data Analytic Methods for Analyzing the Measurement Datasets

Managed by:

Professor K. Matusita
(The Institute of Statistical Mathematics)

Professor Dugué
(the Institute of Statistics
at the University of Paris VI)

Supported by:

JSPS (The Japan Society for the Promotion of Science)

- It was *a memorable event* in the development of data analysis in both countries.

- By the support of the Japan Society for the Promotion of Science (JSPS).

- With only ten or so participants, including Japanese researchers; *Nisihira, Hayakawa, Inagaki, Sugiyama, Sakasegawa, Suzuki, myself, and so on.*

(J-F small meeting, 1978)

- We had a chance to meet *Professor J.-P. Benzécri*.

- We had also a chance to meet *Professors Lebart, Roux, and Jambu*, who were then promising young data analysts of Benzécri's school and France.

(Young Jambu and myself, 1978)

A group of researchers in Japan:

They had achieved considerable advances in data analysis research by developing theories and methods.

The group was led by C. Hayashi,

- The founder of the *quantification methods*,
- The post of General Director of the Institute of Statistical Mathematics
  (ISM; 1974 to 1986)

The concurrent developments in France we knew only partially.

Existence of a method similar to one of *Hayashi's Quantification Methods, Type III* (called the *pattern classification method*)

Meanwhile, as just mentioned, we had a chance to meet Professor Benzécri at the 1978 seminar, and since then Japanese and French researchers have been in a close and lasting relationship with each other.

This was *the beginning of the history* of the mutual development of the new data analysis in France and Japan.

2. The Dawn of Data Analysis

A bridge was built between Japan and France.

Hayashi and some other researchers planned to invite Benzécri to Japan, and *Professor M. Roux* visited Japan on his behalf (as JSPS's invited researcher).

Some seminars and special lectures were planned to introduce the French philosophies of *"analyse des données"* to Japanese researchers:

> *The Institute of Statistical Mathematics (ISM)*
> *The Japanese Classification Society (JCS)*
> *and other places (the universities, and so on)*

The philosophies introduced by Roux was astonishingly new and stimulating. We were very interested in *the ideas of correspondence analysis, automatic classification*, and so on. Roux made an immense contribution by helping us examine how similar the mathematics are *between correspondence analysis and the Type-III Quantification Method.*

We have heated discussions about what statistical analysis was practical, *what statistical science was useful*, and *what data analysis was.*

(Roux and Yajima)

Many notable achievements:

- *Hayashi's quantification methods,*
- *Akaike's information criterion (AIC),*
- *Survey of Japanese National Character, and so on.*

The Institute played an essential part in offering opportunities *to put these new theories into practice*.

It was particularly symbolic that much of the research based on surveys in Japan.

As another case, ...

we invited *Professor L. Lebart*, an authority in data analysis and social survey research, with the support of the JSPS, the CNRS, and the ISM.

(Lebart, 1979)

Lebart and Hayashi had been keeping in close contact with each other. Then, the opportunity was taken to survey *international attitudes to the "Japanese and French national characters,"* both in Japan and France. Lebart visited Japan several times after that. We think highly of him and his colleagues for *their contribution to our survey-based research in Japan*.

(Lebart, 1987)

3. The Foundation of the Japanese Classification Society and Its Relationship to the SFC

As another great contribution to Japan by French researchers:

For the foundation of the *Japanese Classification Society* (JCS) in 1983.

Many societies had been already organized:

- *The SFC had already started in France*
- *The Classification Society of North America (CSNA)*
- *Gesellschaft für Klassifikation (GfKl)*
- *British Classification Society (BCS)*

From *Professor M. Jambu*, we learned a lot of the activities of the classification societies in the USA and Europe.

(Jambu and Tanaka, 1983)

(Jambu, 1987)

Hayashi and some other researchers had obtained additional information, and *decided to found our own society in 1983*.

The JCS is a small society, with a membership of about 200, but it is highly regarded in Japan as the society representing Japan in the international federation.

4. Fruitful Research Interchange Through Later Meetings

Large international conferences held in Japan which presented opportunities to meet many French and other countries' researchers.

For example,

- *The 46th International Statistical Institute Conference (ISI), 1987, Tokyo.*
- *International Biometric Society (IBS), 1984, Tokyo.*

We had a chance to meet many researchers:
Y. Escoufier, Nakache, Bouroche (France), Gower (England), and Rizzi and Lauro (Italy).

(Escoufier, 1987)


(Nakache, 1987)

The Japanese Data Analysts' group appeared in *several large international conferences* in France:

- *International Symposium on "Data Analysis and Informatics"*

  - Second international symposium, October 1979, Versailles (E.Diday, L.Lebart, J.-P.Pages, R.Tomassone, et al).

  - Third international symposium, October 1983, Versailles (E.Diday, M. Jambu, L.Lebart, J.-P.Pages, R.Tomassone).

  - Fourth international symposium, October 1985, Versailles (E.Diday, Y.Escoufier, L.Lebart, J.-P.Pages, Y.Schektman, R.Tomassone).

- *International Conference on "Data Analysis, Learning Symbolic and Numeric Knowledge"*

  - The International Conference, September 1989, Antibes, (E.Diday et al).

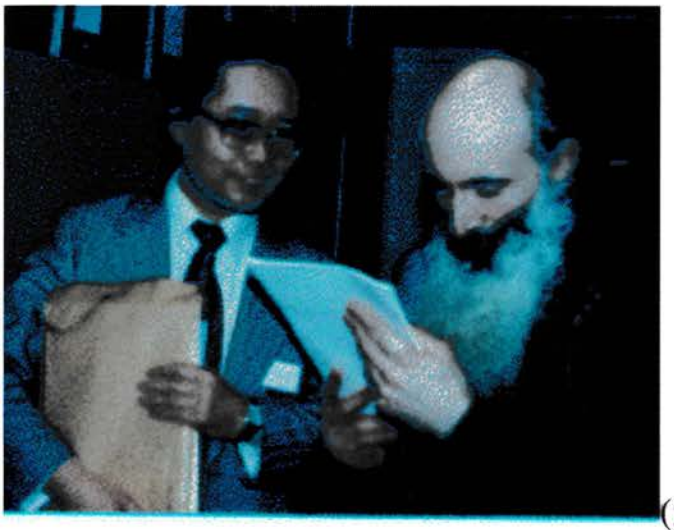
● *At the second meeting in Versailles in 1979:*

C. Hayashi made a speech as an invited speaker and some other Japanese researchers, including me, read their papers.

During their stay in France, *Hayashi and others had met first Professor Benzécri.*

● Other Japanese researchers — *Iwatsubo, Yanai, Ohsumi, Takakura, and Sugiyama* — followed them at the third and fourth meetings, held *in 1983 and 1985*, respectively.

(Hayashi and Benzécri had met first in Paris, 1979)



(in 1979)

(in 1979)

(in 1985)

The second meeting in 1979 must be remembered by Japanese researchers as extremely important. It was *a starting point for close research exchanges* between Japan and many European countries, such as Britain, Italy, Germany, Spain, Switzerland, and so on.

In particuler, *acquaintance with Italian researchers, such as Rizzi and Lauro*, also evolved into a productive relationship.

(in 1979, Paris)

(in 1984, Tokyo)

5. "Analyse des Données" and "Deta Kaiseki"

There are *similarities and differences in the approaches to data analysis* between Japan and France. It is important to emphasize that we agree on the need to develop, through practice, research on the theory and application of data analysis into a new "data science."

As a simple and typical example:

- *Quantification Method, Type-III*
- *Correspondence Analysis*

●Hayashi's Quantification Methods

- Comprise several methods (*Type I to Type VI*)
- *Type III* (*Pattern Classification Method*)
  (in 1952, 10 years earlier than Benzécri's proposal)
- Coincides with *Correspondence Analysis*

●Hayashi's interests:

- Analysis of qualitative data
- The idea of scaling methods, and so on.

●Benzécri's approaches:

- *Correspondence Analysis*
- *Analyse Factorielle des Correspondances*

(appeared about 1962)

●Benzécri and his school succeeded in developing elaborate and varied theories:

- *Correspondence Analysis, and the related methodologies*
- *Multiple correspondence analysis*
- *Automatic classification, etc.*

We saw in many achievements trends emerging which were quite different from those in the United States and Britain.

(Lerman, 1987)

(Diday, 1987)

However, the *"barrier of language"* intervened, and Japanese researchers were unable to gain true recognition for their achievements.

In addition, the *"dialect" or "jargon"* used in research on the *"analyse des données"* made things still worse. Though there has been some improvement, we are still in much the same situation now.

For instance,

- supplementary elements (*éléments supplémentaires*)
- distribution of cloud (*nuage*)
- *analyse factorielle* (not "factor analysis")
- profile (*profils*)
- inertia

simultaneous representation (*représentation simultanée*)

....,

....,

In Japan, because of *various dialects and the uniqueness of Hayashi's theory (very philosophical)*, there was a misunderstanding that *"deta kaiseki"* is very difficult.

The Japanese language used in these papers prevented these achievements from becoming known to overseas researchers.

●However, with publications *by Lebart and Ohsumi in Japanese* (1994), Japanese researchers are now able to get much more information about the research results in France and in other countries.

Differences in language and thought make most Japanese researchers more interested in research in English-speaking countries, which presents a great problem for us to solve.

●Books in English *by Greenacre (1984)* and *Jambu (1983)* are read by many Japanese researchers and students. Those who are interested in *analyse des données* are increasing in number. We do hope researchers in both countries will overcome the barriers, and develop a more active research interchange.

大隅　昇　L.ルバール
A.モリノウ　K.M.ワーウィック　馬場康維

記述的多変量解析法

日科技連

Now, on behalf of many Japanese researchers, we would like to give special thanks to our French colleagues listed here:

Bouroche, Caussinus, Diday,
Durand, Escoufier
the late Escofier, Fichet,
Holmes, Jambu, Lebart, Lerman,
the late Megreditchian,
Morineau, Nakache, Schektman,
Van Cutsem,
and many other French researchers.

We value their efforts and understanding highly.

6. The Expansion of Our Research Interchange Through the Japanese-French Scientific Seminar

Two important results in the history of research interchange between Japan and France:

In the past, *two Japanese-French Scientific Seminars* were arranged.

(1) *At the Institute of Statistical Mathematics, in Tokyo (March 24-26, 1987).*

(2) *The second meeting was in Montpellier, University II in France (August - September 2, 1992).*

●*The first meeting* was organized by Jambu, Hayashi, and Ohsumi:

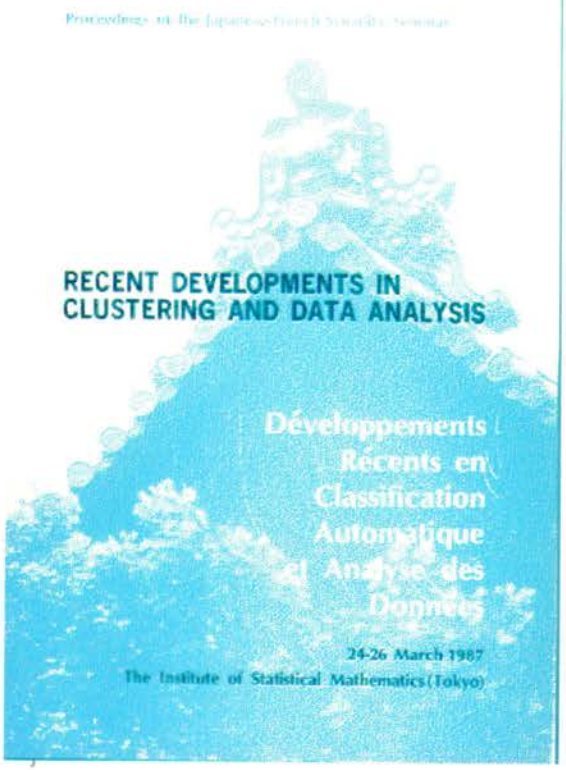
- By support through the Japanese-French Research Interchange Fund from JSPS and CNRS.
- The meeting was held under the auspices of those organizations and ISM.
- Thirty-five papers were presented and the computer and software demonstrations were provided during the seminar.
- This meeting was held in the same year as that of the ISI-87 Conference.

•Over 180 researchers took part in the seminar (much larger than we had expected).

SFC-97/OHP(Section-6) **/// Exhibit 6-3 ///**

•The results can be seen in the book titled:

"Recent Developments in Clustering and Data Analysis — Développements Récents en Classification Automatique et Analyse des Données —"
(Hayashi et al, 1988).



Recent Developments in Clustering and Data Analysis

Développements Récents en Classification Automatique et Analyse des Données

Proceedings of the Japanese-French
Scientific Seminar
March 24–26, 1987

Edited by

Chikio Hayashi
University of the Air
Wakaba, Chiba
Japan

Michel Jambu
Centre National d'Etudes des Télécommunications
Issy Les Moulineaux
France

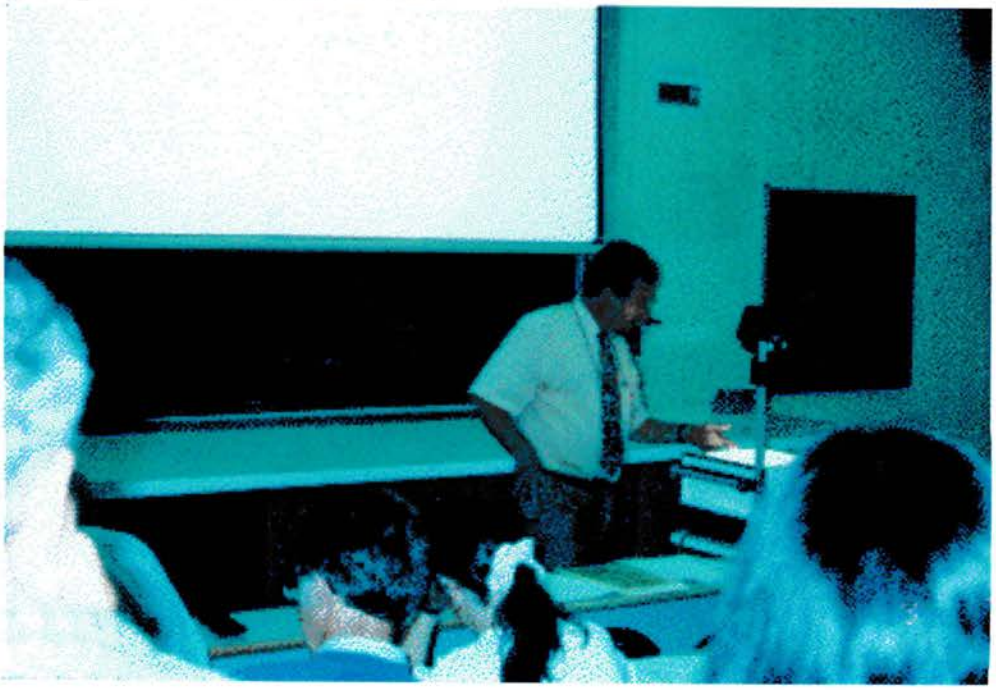
Edwin Diday
INRIA
Domaine de Voluceau
Le Chesnay Cedex
France

Noboru Ohsumi
The Institute of Statistical Mathematics
Minato-ku, Tokyo
Japan

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
Berkeley London Sydney
Tokyo Toronto

● *The second meeting* was organized by two members of the Escoufier group, Hayashi and Ohsumi.

(Prof. Diday's presentation, 1992)



(Prof. Fichet's presentation, 1992)

•The term *"Data Science"* appeared in this meeting for the first time. This was a landmark in the history of data analysis studies.

•After the scientific seminar, Escoufier, Hayashi, and Ohsumi were engaged as editors of the proceedings, and while writing the preface of the book we used the term *"Data Science."*

(*) I remember that was when we made some arrangements at ICOT-4 held at Morocco in 1994.

The book was published under the title of:

"Data Science and Its Applications
—La Science des Données
et ses Applications—"
(Escoufier et al., 1995).

Data Science and Its Applications

-La science des données et ses applications-

Edited by
Yves Escoufier
Bernard Fichet
Edwin Diday
Ludovic Lebart
Chikio Hayashi
Noboru Ohsumi
Yasumasa Baba

ACADEMIC PRESS, INC.
Harcourt Brace & Company
Tokyo San Diego New York Berkeley
Boston London Sydney Toronto

7. Relationship to IFCS — Changing from a Linear to a Spatial Perspective

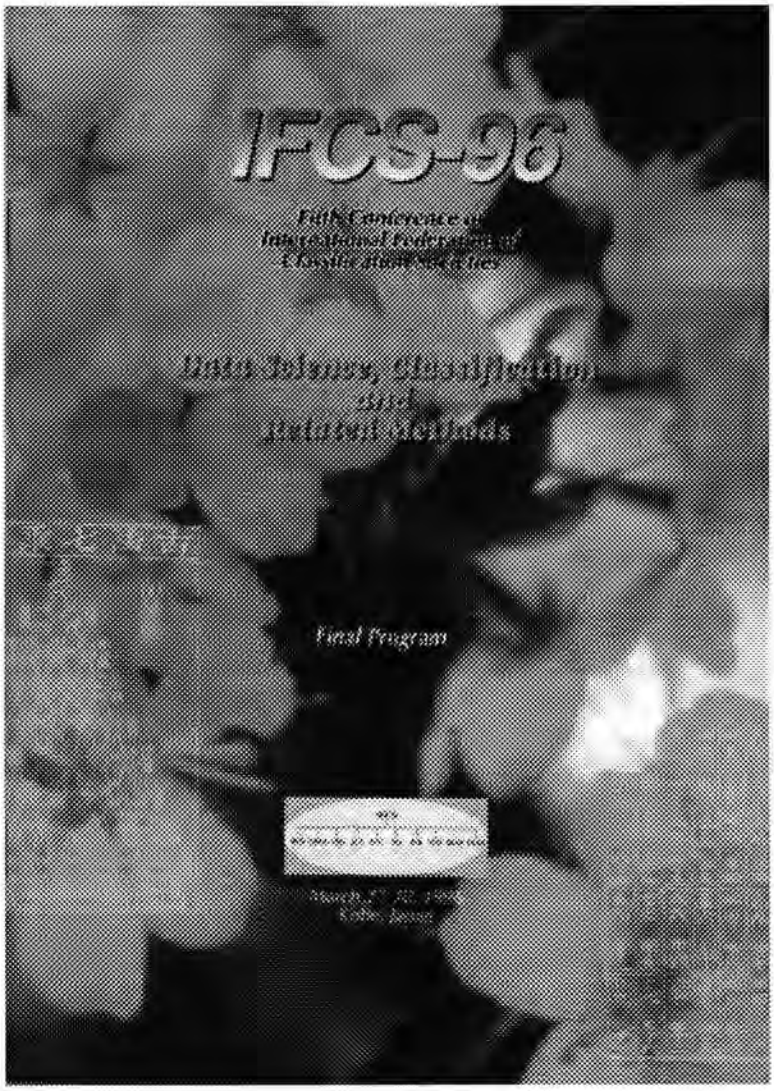
Japan's foreign relations in this field began with the research interchange between Japan and France. The relation was at first a linear one, so to speak, but more extensive relations developed afterwards, one of the results of which was the foundation of the IFCS, and another was the actual exchanges with many other countries.

Thanks to the efforts of all concerned, the IFCS *was founded in 1985* to federate the classification societies from many countries.

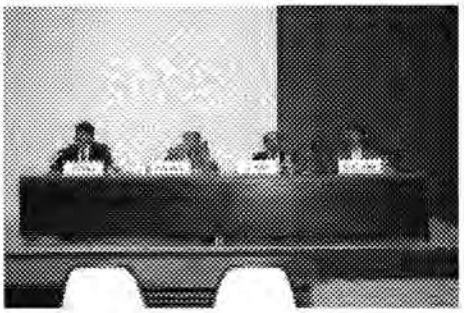
H. Bock's devotion to the First IFCS International Conference held at Aachen in Germany in 1987 deserves special mention. It was the first meeting held by the federation of BCS, CSNA, GfKl, JCS, SFC, and SIS. After that, IFCS meetings were held *in Virginia (1989), Edinburgh (1991), Paris (1993), and Kobe (1996).*

Fifth IFCS-96 Conference in March 1996

So far, twelve societies have joined the federation. It may be very hard to organize it well, and the number of federated societies is small, but its aim is positive and constructive. Japan was in the fortunate position of being able *to host the Fifth IFCS-96 Conference* in March 1996, after the fourth in France in 1993.



This is certainly one of the greatest results of *the twenty years' research interchange between Japan and France.* Thus, the association between both countries may have undergone a marked *change from a linear to the spatial relationship.*



(Panel session)

(President's address)

(Reception)

(Welcome party)

8. Toward Data Science
— Prospects in Data Analysis

Generally speaking, it has been a long time now since attention was drawn to a decline and crisis in statistical science.

Nevertheless, no marked improvement has been made. We find it very surprising and odd that *no university in either France or Japan has a department of statistics.*

Most of the students of statistical analysis or data analysis are enrolled in such departments *as information science, biometrics, psychology, and the like*, and there they are engaged in their studies.

In Japan, in the field of statistical science or data science, we have *only one specialized research institute, ISM*, and as for graduate schools, we have only one statistical science course, also at the ISM.

In recent times in Japan, there has been a great deal of discussion over the guidelines for scientific research. Especially, in the fields of computational science and informatics, many thought it necessary to look over *how to advance the research.*

In the course of this review process, a lot of research projects overseas are being introduced for the purpose of comparison or benchmarking.

Models drawn from *large-scale national research centers, such as INRIA or the organization of CNRS* have drawn much interest.

In the field of data analysis, a large-scale institute of

computational science or informatics

is planned *as part of the structural reorganization program in Japan.* We are hoping that this new development will come to fruition, and, at the same time, we are looking to develop a new system.

At the moment, however, no definite plan or idea is to be found.

It is necessary for us to decide *in what direction we should progress*.

We must re-examine our attitude toward our research and make ourselves refreshed.

For that purpose, we might *have to seek collaboration and cooperation* with other fields, or even consider the possibility of organization and integration.

We might *have to abandon such terms as statistical science or data analysis*, and choose, for example, *"Data Science"* as a new keyword and concept. We believe that such a concept can help to guide and foster a fruitful and expanding relationship between both countries in the future.

We very much hope *this new age of "Data Science"* will come to fruition, and that what we have done in the history of data analysis research will be of enduring benefit to the coming science and to future research.

Acknowledgements

I would like to express my grateful acknowledgements to all staff and members of the Organization Committee of *the 20th Anniversary of SFC* for giving me the opportunity to present this report, and to all the Japanese and French researchers for making great efforts toward the development of Data Science.