

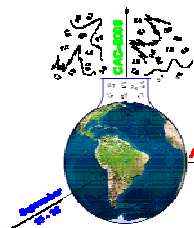
10th International Conference on Chemometrics in Analytical Chemistry
CAC-2006

CHEMOMETRICS IN THE TROPICS

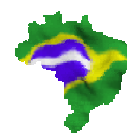
Nature, Medicine and Industry

September 10-15

Águas de Lindóia, SP, BRAZIL



Águas de Lindóia
SP - BRAZIL



Data Mining Session, OP27

Literature and Internet Database Mining in a Study About the Word CHEMOMETRICS

Rudolf Kiralj and Márcia M. C. Ferreira



Laboratório de Quimiometria Teórica e Aplicada (LQTA)



Instituto de Química

Universidade Estadual de Campinas (UNICAMP)

Campinas, SP, 13083-970, BRAZIL



E-mails: rudolf@iqm.unicamp.br, marcia@iqm.unicamp.br

URL: <http://lqta.iqm.unicamp.br>

Keywords

- bibliometrics (WOS=Web of Science, SCI=Science Citation Index-Expanded) of CHEMOMETRICS
- webometrics (Google, Yahoo) of CHEMOMETRICS
- linguistics of CHEMOMETRICS
- chemometrics-development relationships

Studied aspects about the word CHEMOMETRICS:

- origin
- history
- writing and pronunciation in different languages
- relations between the found languages
- qualitative/quantitative parameters for chemometric activity
- parameters of past, present and future trends in chemometrics

Motivation

1996-1999: Why **chemometrics** and how to say it in my language?

2002: the first study in the LQTA – **chemometrics** in 16 languages

<http://pcserver.iqm.unicamp.br/~rudolf/chemometrics.html>

2003 and after: Dr. K. Faber's study – **chemometrics** in 30 languages and relationship **chemometrics-chemometry**

February 2006: Dr. K. Faber's **chemometrics-chemometry** at ICS-L

Other online **chemometrics-chemometry** divisions and discussions:

-in German: <http://www.pharmazie.uni-wuerzburg.de/AKBaumann/chemometrik.html>

-in Russian: <http://rcs.chph.ras.ru/rcsin.htm>

-in Croatian: <http://www.pbf.hr/hr/layout/set/print/content/view/sitemap/2>

-in Macedonian: <http://hemija.net/statii/statija.php?ids=104>

Methods

-database minings in the WOS and SCI

-Google and Yahoo searches and internet surfings

-use of diverse literature (in electronic and printed forms)

-generation of bibliometric and webometric descriptors or indices

-selection of country development indices (from literature)

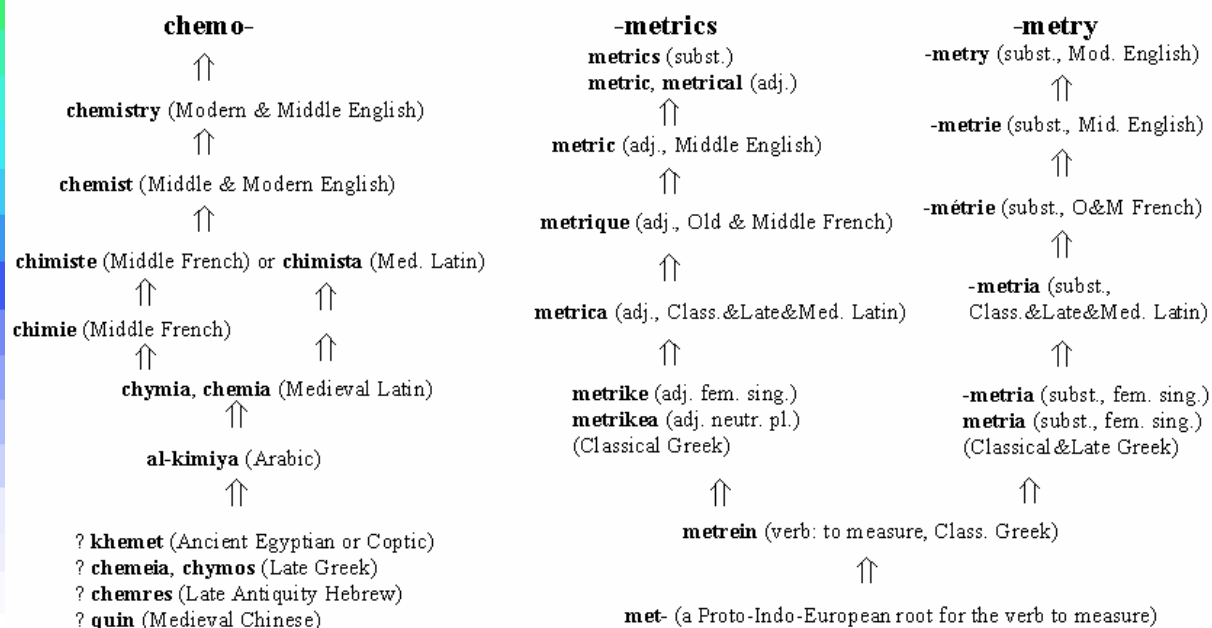
-data analysis: simple statistics and chemometrics-development relationships (exploratory analysis and PLS regression models)

CHEMOMETRICS: total etymology and metrics/metry distinction

chemometrics = chemo- + -metrics

(chemometry = chemo- + -metry)

P R E S E N T D A Y



ANTIQUITY OR NEOLITHIC

NEOLITHIC

CHEMOMETRICS: early history and evolution

Date	Event
1971	Prof. Svante Wold: kemometri and chemometrics
1972	Prof. S. Wold: <u>Forskningsgruppen för Kemometri</u> or <u>Kemometrigruppen</u> ; and publishes 1st article with kemometri
1973	1st WOS article with chemometrics in the name of a research group (Prof. S. Wold's)
1974	Prof. S. Wold: the definition of chemometrics in analogy with that of biometrics, econometrics and other -metrics sciences
1974	Professors S. Wold and Bruce R. Kowalski: The <u>International Chemometrics Society</u>
1975	Prof. B. R. Kowalski: 1st first article with chemometrics in the title, and chemometrics defined as a new chemical discipline
1976	1st <u>Chemometrics Newsletter</u> (Prof. B. R. Kowalski)
1976	1st <u>Symposium "Chemometrics: Theory and Applications"</u> and 1st book on chemometrics (Editor: Prof. B. R. Kowalski)
1978	<u>1st International Conference on Chemometrics in Analytical Chemistry (CAC)</u>
1979	1st time chemometric in the WOS
1980	1st review on Chemometrics in the Fundamental Reviews of Analytical Chemistry
1981	1st chemometricians in the WOS
1984	1st chemometrics in WOS publications from Asia
1986	<u>Chemometrics and Intelligent Laboratory Systems</u>
1987	<u>Journal of Chemometrics</u>
1987	1st <u>Colloquium Chemiometricum Mediterraneum</u>
1988	1st <u>professorship in chemometrics</u> (Prof. S. Wold)
1988	1st <u>Scandinavian Symposium on Chemometrics</u>

CHEMOMETRICS: linguistic reality → POSTER

CHEMOMETRICS was found worldwide in:

- 48 languages
- 10 writing systems
- 82 orthographic forms
- 127 standard pronunciation forms

and on 6 continents:

North and South America: 4 languages

Africa: 1 language

Australia: 1 language

Asia: 13 languages

Europe: 34 languages

Orthographic forms are characterized by:

- end form types (-TRIX)
- relative frequency
- geographic distribution and preference

Table 1. Linguistic of the word chemometrics in standard national languages

Lang.	Language	Abbr. (writing)	Latin transcription ¹	No. First/Last	Class ²	Freq ³
en-1	English	chemometrics	chemometrics	1	1	100
en-2	English	chemometry	chemometry	1	1	100
en-3	English	chemiometrics	chemiometrics	1	1	100
en-4	English	chemimetry	chemimetry	1	1	100
en-5	English	chemometrics	chemometrics	1	1	100
en-6	English	chemometrics	chemometrics	1	1	100
en-7	English	chemometrics	chemometrics	1	1	100
en-8	English	chemometrics	chemometrics	1	1	100
en-9	English	chemometrics	chemometrics	1	1	100
en-10	English	chemometrics	chemometrics	1	1	100
en-11	English	chemometrics	chemometrics	1	1	100
en-12	English	chemometrics	chemometrics	1	1	100
en-13	English	chemometrics	chemometrics	1	1	100
en-14	English	chemometrics	chemometrics	1	1	100
en-15	English	chemometrics	chemometrics	1	1	100
en-16	English	chemometrics	chemometrics	1	1	100
en-17	English	chemometrics	chemometrics	1	1	100
en-18	English	chemometrics	chemometrics	1	1	100
en-19	English	chemometrics	chemometrics	1	1	100
en-20	English	chemometrics	chemometrics	1	1	100
en-21	English	chemometrics	chemometrics	1	1	100
en-22	English	chemometrics	chemometrics	1	1	100
en-23	English	chemometrics	chemometrics	1	1	100
en-24	English	chemometrics	chemometrics	1	1	100
en-25	English	chemometrics	chemometrics	1	1	100
en-26	English	chemometrics	chemometrics	1	1	100
en-27	English	chemometrics	chemometrics	1	1	100
en-28	English	chemometrics	chemometrics	1	1	100
en-29	English	chemometrics	chemometrics	1	1	100
en-30	English	chemometrics	chemometrics	1	1	100
en-31	English	chemometrics	chemometrics	1	1	100
en-32	English	chemometrics	chemometrics	1	1	100
en-33	English	chemometrics	chemometrics	1	1	100
en-34	English	chemometrics	chemometrics	1	1	100
en-35	English	chemometrics	chemometrics	1	1	100
en-36	English	chemometrics	chemometrics	1	1	100
en-37	English	chemometrics	chemometrics	1	1	100
en-38	English	chemometrics	chemometrics	1	1	100
en-39	English	chemometrics	chemometrics	1	1	100
en-40	English	chemometrics	chemometrics	1	1	100
en-41	English	chemometrics	chemometrics	1	1	100
en-42	English	chemometrics	chemometrics	1	1	100
en-43	English	chemometrics	chemometrics	1	1	100
en-44	English	chemometrics	chemometrics	1	1	100
en-45	English	chemometrics	chemometrics	1	1	100
en-46	English	chemometrics	chemometrics	1	1	100
en-47	English	chemometrics	chemometrics	1	1	100
en-48	English	chemometrics	chemometrics	1	1	100
en-49	English	chemometrics	chemometrics	1	1	100
en-50	English	chemometrics	chemometrics	1	1	100
en-51	English	chemometrics	chemometrics	1	1	100
en-52	English	chemometrics	chemometrics	1	1	100
en-53	English	chemometrics	chemometrics	1	1	100
en-54	English	chemometrics	chemometrics	1	1	100
en-55	English	chemometrics	chemometrics	1	1	100
en-56	English	chemometrics	chemometrics	1	1	100
en-57	English	chemometrics	chemometrics	1	1	100
en-58	English	chemometrics	chemometrics	1	1	100
en-59	English	chemometrics	chemometrics	1	1	100
en-60	English	chemometrics	chemometrics	1	1	100
en-61	English	chemometrics	chemometrics	1	1	100
en-62	English	chemometrics	chemometrics	1	1	100
en-63	English	chemometrics	chemometrics	1	1	100
en-64	English	chemometrics	chemometrics	1	1	100
en-65	English	chemometrics	chemometrics	1	1	100
en-66	English	chemometrics	chemometrics	1	1	100
en-67	English	chemometrics	chemometrics	1	1	100
en-68	English	chemometrics	chemometrics	1	1	100
en-69	English	chemometrics	chemometrics	1	1	100
en-70	English	chemometrics	chemometrics	1	1	100
en-71	English	chemometrics	chemometrics	1	1	100
en-72	English	chemometrics	chemometrics	1	1	100
en-73	English	chemometrics	chemometrics	1	1	100
en-74	English	chemometrics	chemometrics	1	1	100
en-75	English	chemometrics	chemometrics	1	1	100
en-76	English	chemometrics	chemometrics	1	1	100
en-77	English	chemometrics	chemometrics	1	1	100
en-78	English	chemometrics	chemometrics	1	1	100
en-79	English	chemometrics	chemometrics	1	1	100
en-80	English	chemometrics	chemometrics	1	1	100
en-81	English	chemometrics	chemometrics	1	1	100
en-82	English	chemometrics	chemometrics	1	1	100
en-83	English	chemometrics	chemometrics	1	1	100
en-84	English	chemometrics	chemometrics	1	1	100
en-85	English	chemometrics	chemometrics	1	1	100
en-86	English	chemometrics	chemometrics	1	1	100
en-87	English	chemometrics	chemometrics	1	1	100
en-88	English	chemometrics	chemometrics	1	1	100
en-89	English	chemometrics	chemometrics	1	1	100
en-90	English	chemometrics	chemometrics	1	1	100
en-91	English	chemometrics	chemometrics	1	1	100
en-92	English	chemometrics	chemometrics	1	1	100
en-93	English	chemometrics	chemometrics	1	1	100
en-94	English	chemometrics	chemometrics	1	1	100
en-95	English	chemometrics	chemometrics	1	1	100
en-96	English	chemometrics	chemometrics	1	1	100
en-97	English	chemometrics	chemometrics	1	1	100
en-98	English	chemometrics	chemometrics	1	1	100
en-99	English	chemometrics	chemometrics	1	1	100
en-100	English	chemometrics	chemometrics	1	1	100

Orthographic variants (forms) or typo mistakes? Scientific convention or freedom of choice?

6 English forms:	construction	freq.
standard	CHEMOMETRICS → CHEMO- + -METRICS	(>99%)
alternative?	CHEMOMETRY → CHEMO- + -METRY	(<0.5-10%)
typo?	CHEMIOMETRICS → CHEMIO- + -METRICS	(<0.5%)
typo?	CHEMIOMETRY → CHEMIO- + -METRY	(<0.5%)
typo?	CHEMIMETRICS → CHEMI- + -METRICS	(<0.5%)
typo?	CHEMIMETRY → CHEMI- + -METRY	(<0.5%)

Obvious typos: CHEMMETRICS, HEMOMETRICS, CHEMEOMETRICS, CHEMEMETRICS...

Native English speakers:
 -METRICS → application of statistics and mathematics to a field of study
 -METRY → process or science of measuring in a field of study

Some other examples:

Afrikaans:

CHEMOMETRIE → CHEMO- + -METRIE (60-90%)

CHEMOMETRIKE → CHEMO- + -METRIKA (10-40%)

Croatian:

KEMOMETRIJA → KEMO- + -METRIJA (53-60%)

KEMOMETRIKA → KEMO- + -METRIKA (40-47%)

German:

CHEMOMETRIE → CHEMO- + -METRIE (90-99%)

CHEMOMETRIK → CHEMO- + -METRIK (0.5-10%)

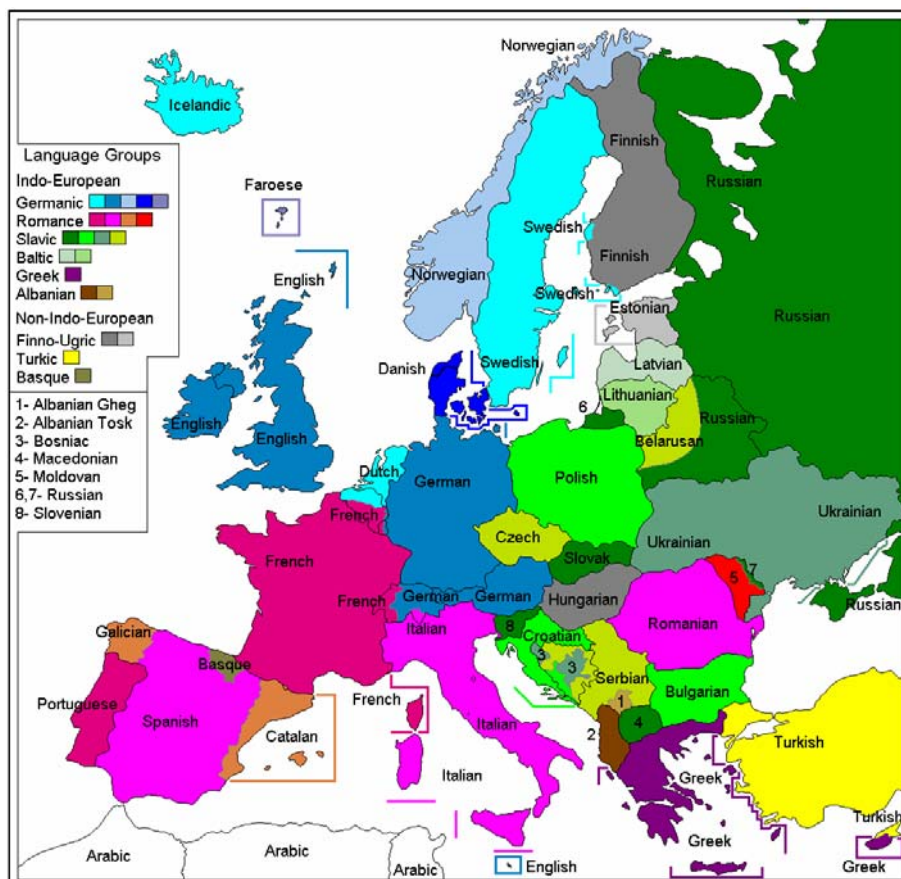
Indonesian:

KEMOMETRI → KEMO- + -METRI (47-53%)

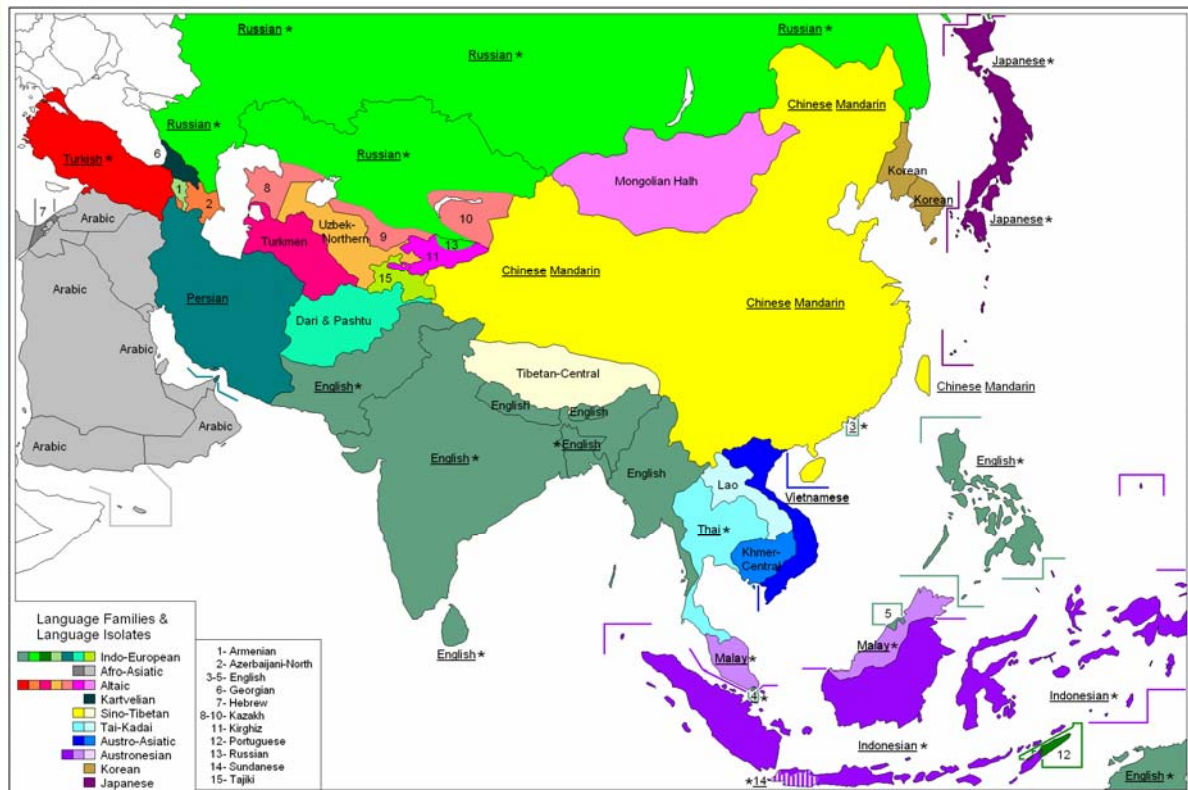
KEMOMETRIK → KEMO- + -METRIK (40-47%)

KEMOMETRIKA → KEMO- + -METRIKA (0.5-10%)

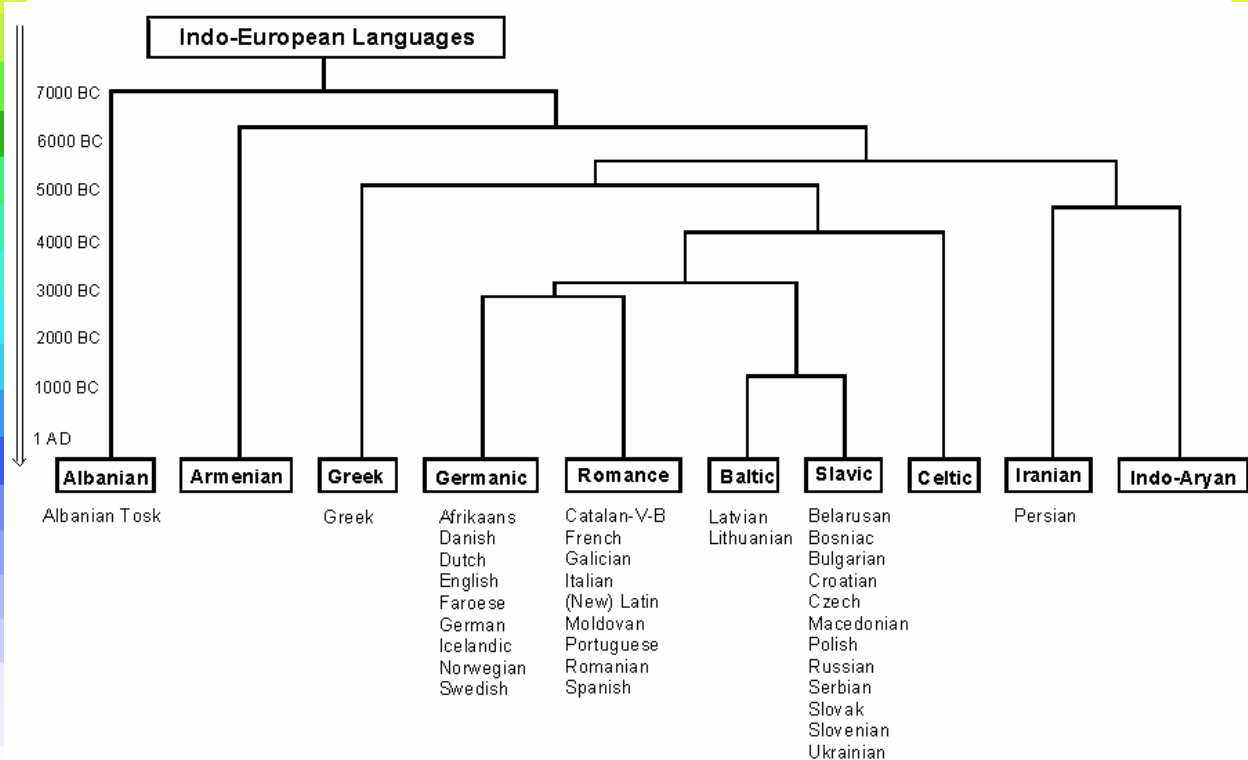
Europe: linguistic situation in science and higher education



Asia: linguistic situation in science and higher education

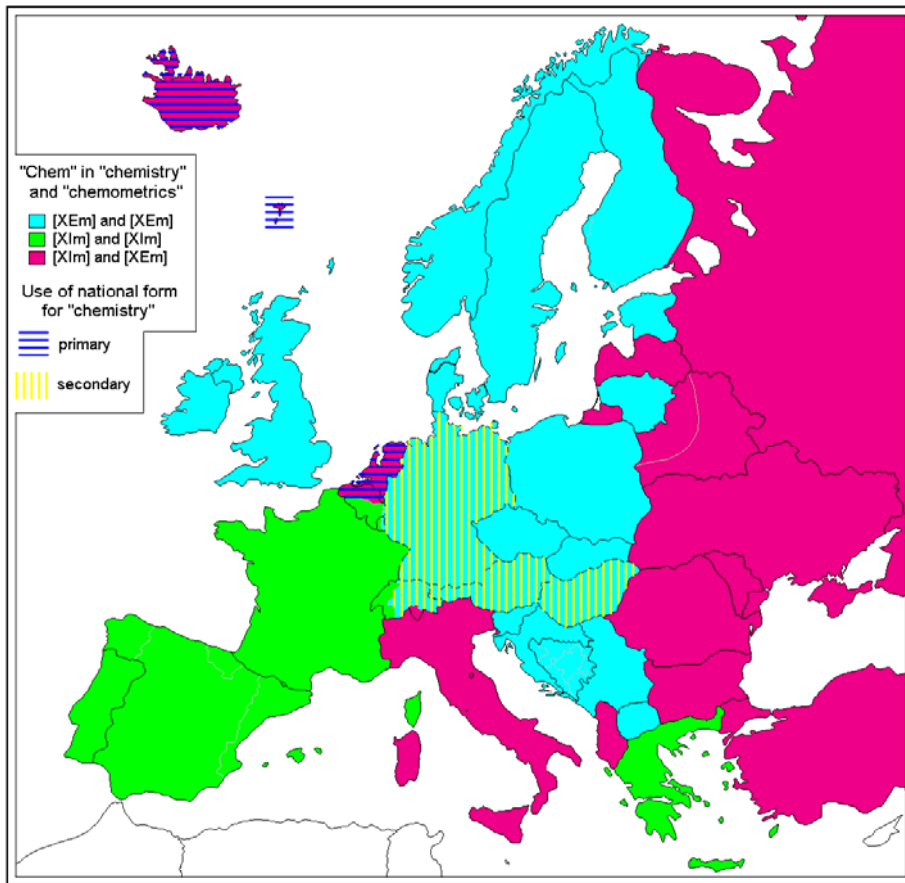


Indo-European family of languages and its living branches

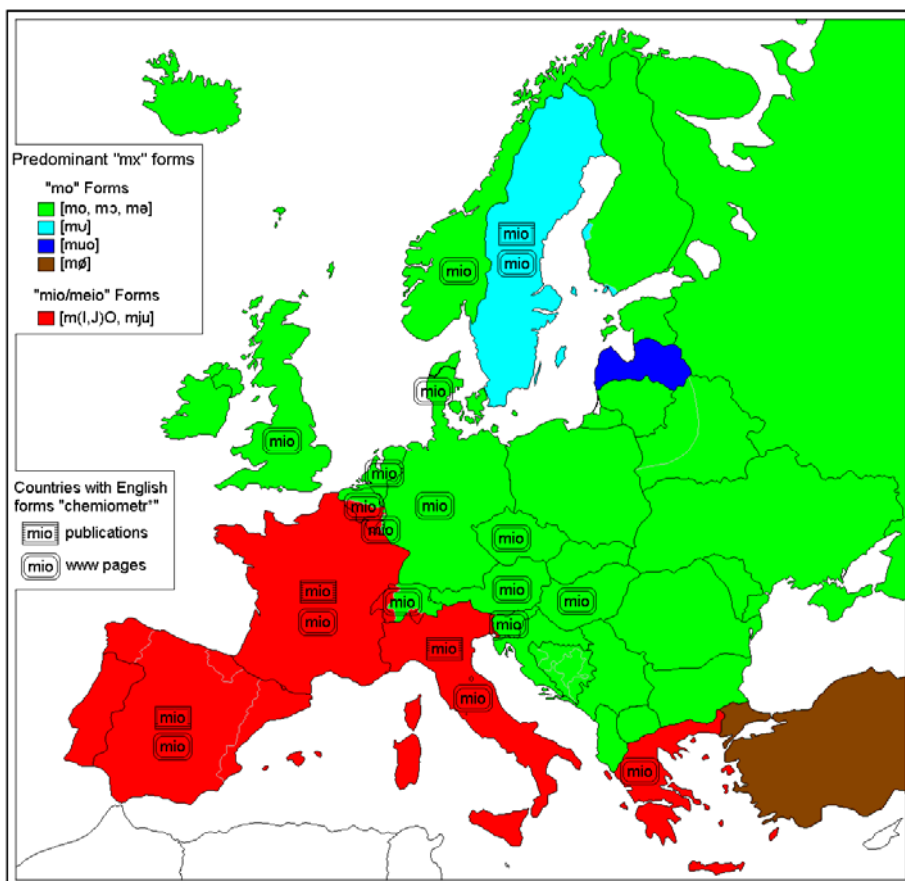


Lexicostatistical dendrogram adapted from L. L. Cavalli-Sforza: *Genes, Povos e Línguas*, Companhia das Letras, São Paulo, SP, 2000, p. 215.

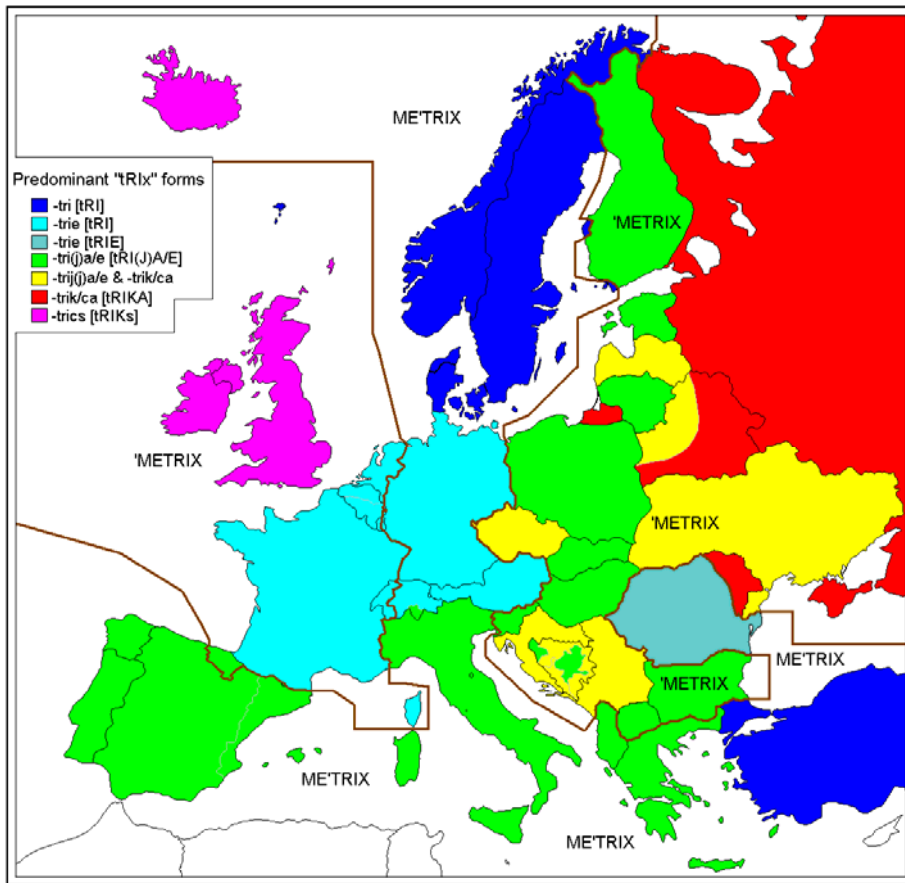
Europe: CHEM- in "chemometrics" and "chemistry"



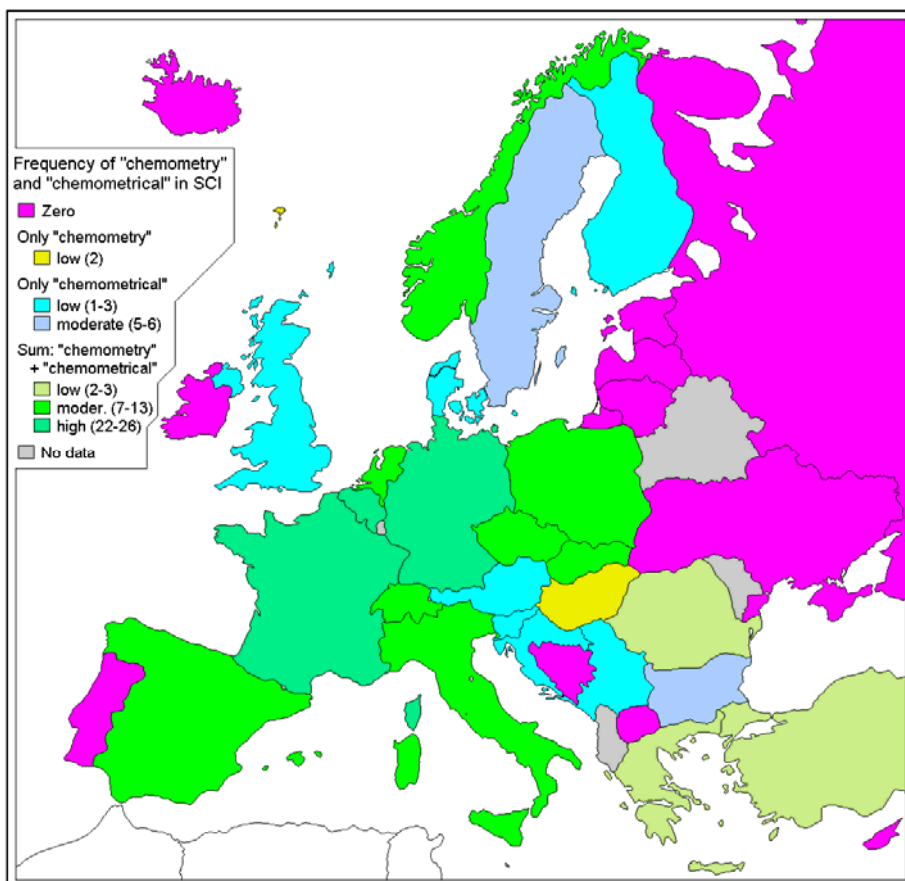
Europe: -MO-/MIO- in "chemometrics"



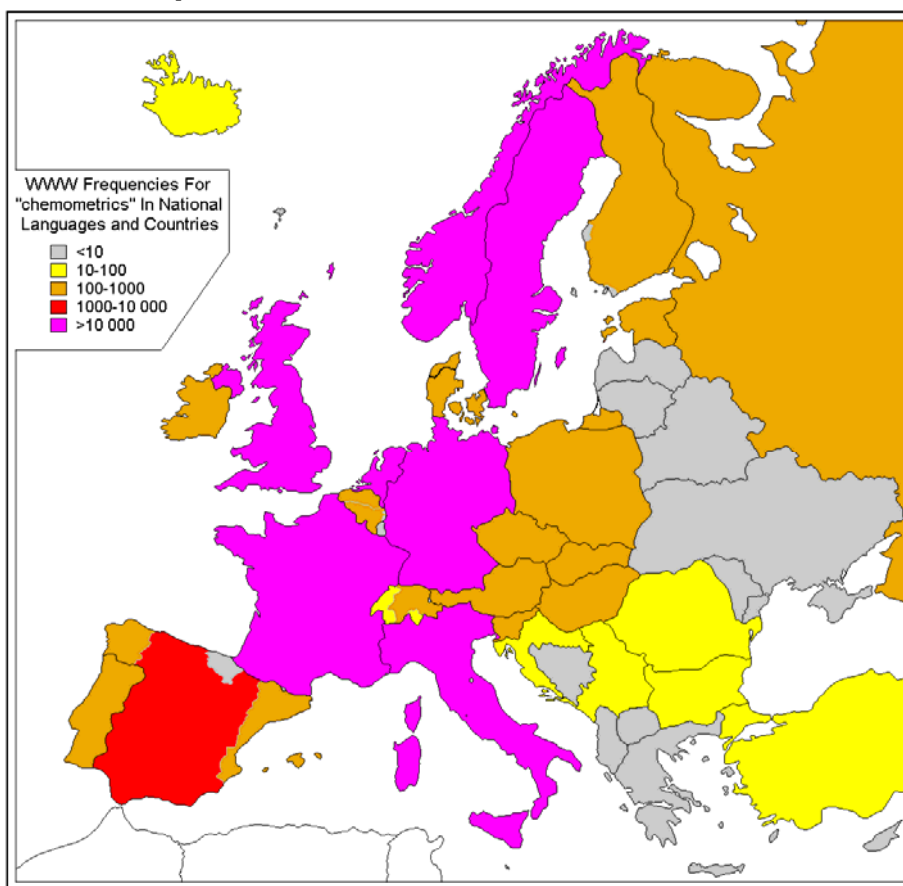
Europe: -TRIX in "chemometrics"



Europe: "chemometry" and "chemometrical"



Europe: webometrics of “chemometrics”

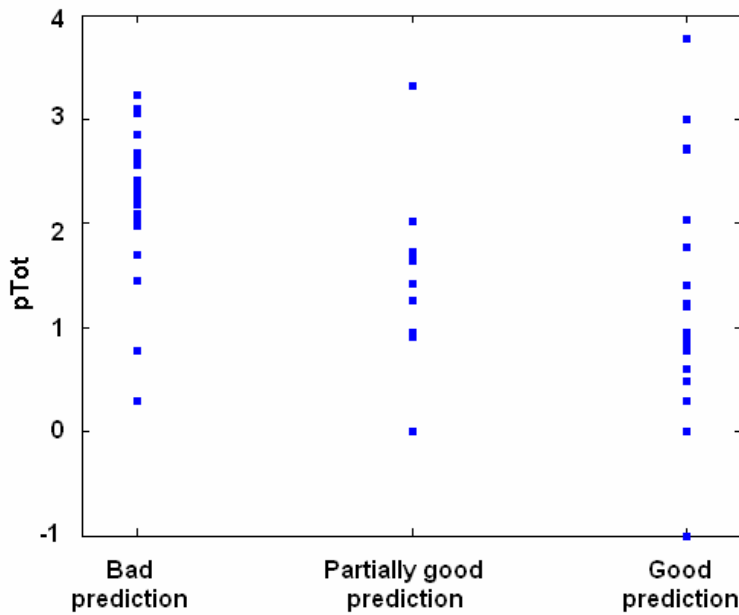


CHEMOMETRICS: orthographic and pronunciation pluralism

Five mechanisms:

- 1) Etymological → **K** or **I,J** end forms **-TRIX** → Class. Gr. Adj./Sub.
- 2) International scientific collaboration → countries with modest scientific production may lay in foreign influences: linguistic & genetic ties; geographic proximity; traditional historical, cultural, economic, scientific and political relationships
- 3) Languages covering large territories and populations → there are more language standards and regions with different linguistic preference
- 4) Countries and political entities speaking the same language → linguistic diversity
- 5) English as the universal language of science → built by native and non-native speakers working in science

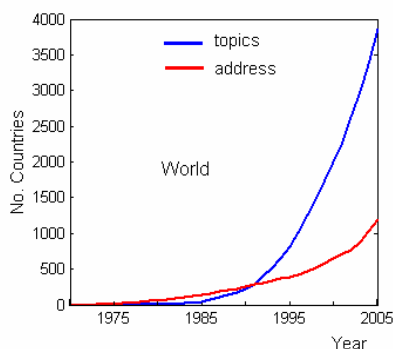
CHEMOMETRICS: prediction of K/I,J end forms –TRIX based on international scientific collaboration



$pTot = \log(Tot)$
 Tot – total No. scientific publications of a country in the SCI (1945/1954-2005)

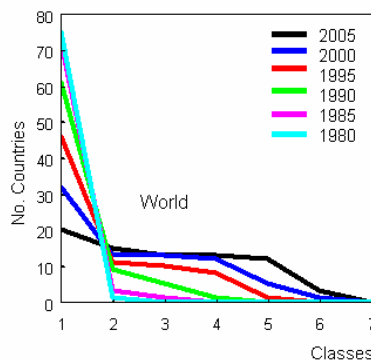
Prediction: -TRIK OR -TRI/TR(I)JA end forms for a language and country depending on % scientific publications done in collaboration with countries that use predominantly either –TRIK or –TRI/TR(I)JA

CHEMOMETRICS: some past, present and future trends



Increasing trend of No. SCI publications with “chemometr*” in topics (Pub) and address

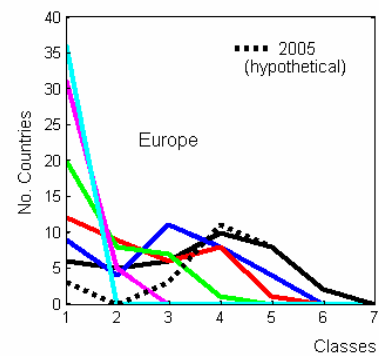
Normal curve within:
 -Europe: 10 years
 -World: 15 years
 -World-total: 70 years



Distribution function for Pub. Classes belong to log units: 1 (0-0.5 units), 2 (0.5-1), 3 (1-1.5), 4 (1.5-2), 5 (2-2.5), 6 (2.5-3), and 7 (3-3.5). Hypothetical Europe: USSR, Czechoslovakia and Yugoslavia.

The tendency of normal curve formation is visible, especially in Europe.

Eastern Europe political changes slow down this trend.

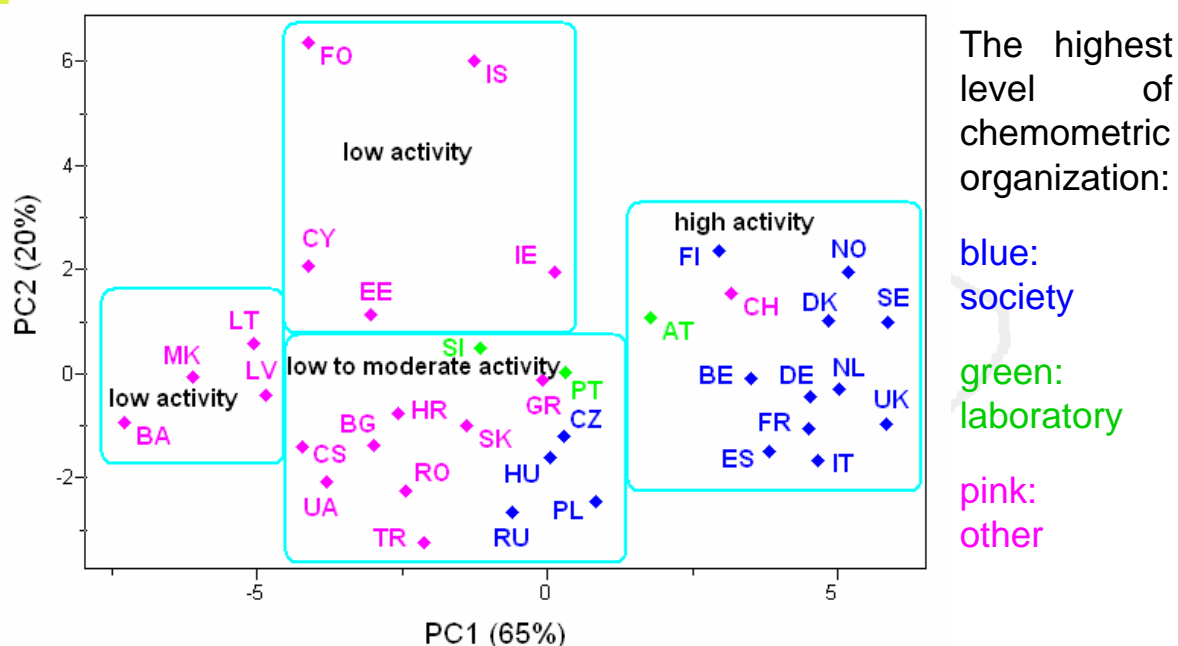


Bibliometric, webometric and country development indices

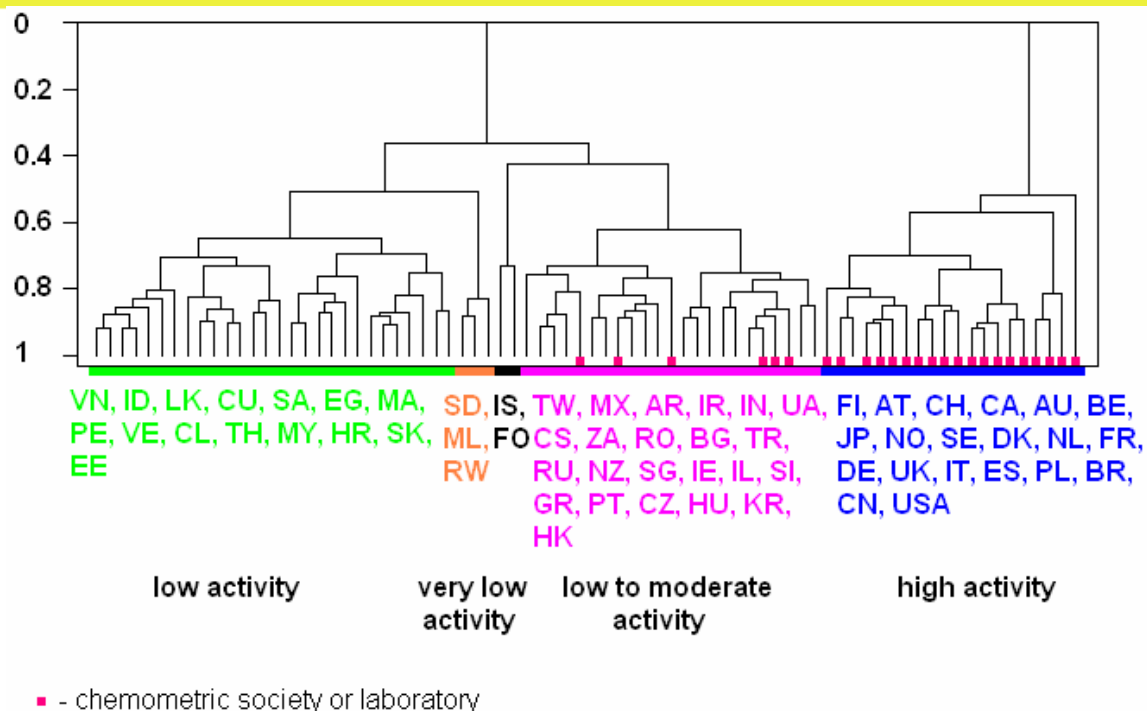
Descriptor	Definition	Method	Nature*
pPub	$pPub = \log(Pub)$, Pub – No. publications with “chemometr*” in topics (title, keywords, abstract)	SCI	B
1stpubl	1st publication date related to Pub	SCI	B
pDNC	$pDNC = \log(DNC)$. DNC – annual No. publications for Pub in the last 5 years	SCI	B
pTot	$pTot = \log(Tot)$. Tot – total No. scientific publications	SCI	B
1stTot	1st publication date related to Tot	SCI	B
pJCpubs	$pJCpubs = \log(JCpubs)$, JCpubs – No. publications in <i>J. Chemometr.</i>	SCI	B
pCILSpubs	$pCILSpubs = \log(CILSpubs)$, CILSpubs – No. publications in <i>Chemometr. Intell. Lab. Syst.</i>	SCI	B
pChempubs	$pChempubs = \log(Chempubs)$, Chempubs – No. publications any of the two journals	SCI	B
pWWW	$pWWW = \log(WWW+1)$, WWW – No. hits for “chemometrics”	Google	W
pJCwww	$pJCwww = \log(JCwww)$, JCwww – No. hits for “Journal of Chemometrics”	Google	W
pCILSwww	$pCILSwww = \log(CILSwww)$, CILSwww – No. hits for “Chemometrics and Intelligent Laboratory Systems”	Google	W
pChJwww	$pChJwww = \log(ChJwww)$, ChJwww – No. hits for any of the two journals	Google	W
pCh1www	$pCh1www = \log(Ch1www)$, Ch1www – No. hits for “chemometrics” and its 7 derivatives	Google	W
pCh2www	$pCh2www = \log(Ch2www)$, Ch2www – No. hits for “chemistry” and its 6 derivatives	Google	W
GDP	Gross Domestic Product per capita in US\$ (2004)	Liter.	D
RIRD	No. researchers in research and development per million people (1999-2001)	Liter.	D
HDI	Human Development Index (2002)	Liter.	D
connec	Connectivity – physical infrastructure for the information and communication technology (2000)	Liter.	D
access	Access – wider determinants of access to the information and communication technology (2000)	Liter.	D
policy	Policy environment of the information and communication technology (2001-2002)	Liter.	D
diffusion	Diffusion of the information and communication technology (2000)	Liter.	D
DAI	Digital Access Index, the overall ability to access and use information and communication technology (2002)	Liter.	D

*Nature: B –bibliometric, W – webometric, D – development index.

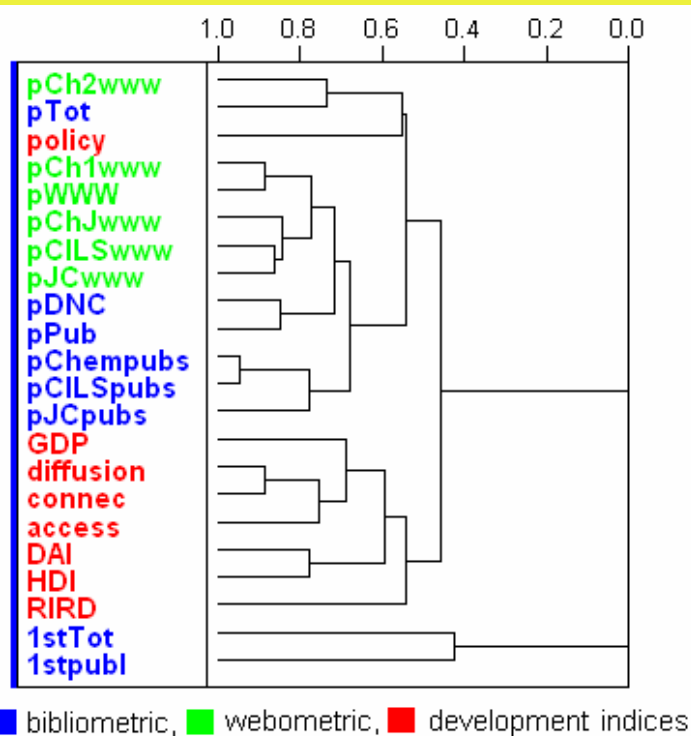
CHEMOMETRICS-DEVELOPMENT RELATIONSHIPS



PCA for Europe based on the 22 descriptors. General pattern of chemometric, chemical and scientific publishing in the WOS-SCI and online: high, low to moderate and low activity. World data show extension of these trends (not presented).



HCA for the world based on the 22 descriptors. General pattern of chemometric, chemical and scientific publishing in the WOS-SCI and online: high, low to moderate, low and very low activity. Europe data show to be a subset of these trends (not presented).



HCA dendrogram with the 22 descriptors for the world, showing noticeable correlations between the development (country development) and bibliometric/webometric descriptors (chemometric activity).

QUANTITATIVE CHEMOMETRICS-DEVELOPMENT RELATIONSHIPS

Prediction of bibliometric and webometric indices using the 8 development indices

Representative examples

$p_{Pub} = \log(Pub)$ for the world

PLS model: $Q = 0.741$, $R = 0.774$, $SEV = 0.551$, $SEP = 0.526$, 74 samples, 2PCs (86%)

$p_{Chempubs} = \log(Chempubs)$ for Europe

Chempubs – No. publications in *J. Chemometr. & Chemometr. Intell. Lab. Syst.* published by a country

PLS model: $Q = 0.811$, $R = 0.833$, $SEV = 0.437$, $SEP = 0.425$, 34 samples, 1PC (84%)

$p_{WWW} = \log(WWW+1)$

WWW – No. Google hits for CHEMOMETRICS for a country domain

PLS model: $Q = 0.774$, $R = 0.810$, $SEV = 0.744$, $SEP = 0.702$, 74 samples, 2PCs (85%)

CHEMOMETRICS: CONCLUSIONS

The word CHEMOMETRICS:

- exists in many languages, mostly as **chemx-** + **-metrix**
- is defined by many factors in a language and country: linguistics & genetics, geography and history, international scientific collaborations
- may serve to generate chemometric activity descriptors in order to see: 1) the trends in chemometrics along time; 2) characterize chemometric activities worldwide; 3) correlate these descriptors with country development indices

THERE ARE VISIBLE QUALITATIVE AND EVEN QUANTITATIVE CORRELATIONS BETWEEN CHEMOMETRICS AND COUNTRY DEVELOPMENT DUE TO SCIENTIFIC AND TECHNOLOGICAL DEVELOPMENT.

THERE ARE OTHER FACTORS WHICH ALSO DETERMINE CHEMOMETRIC ACTIVITY OF A COUNTRY.