JOURNAL OF CHEMOMETRICS
J. Chemometrics 2006; **20**: 247–272
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/cem.1001

The past, present, and future of chemometrics worldwide: some etymological, linguistic, and bibliometric investigations†

R. Kiralj and Márcia M. C. Ferreira*

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas 13083-970, SP, Brazil

Received 26 July 2006; Revised 28 September 2006; Accepted 1 October 2006

Internet surfing for the word chemometrics in national languages and, in the Science Citation Index (SCI), searching for articles containing chemometr* were performed. The bibliometric, webometric, and country development descriptors from literature were then treated by Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). In total, 82 written and 127 pronunciation forms of chemometrics were found in 48 languages worldwide. The forms ending in '*-y*' (*chemometry*) and '*-ics*' (*chemometrics*) can be grouped into at least three groups (I, J, K). Scientific collaboration, country development, geography, history, and language were shown to be important determinants in creation of form(s) of chemometrics in a particular country or language. PCA and HCA show that tradition in chemometrics, level of country development, and its scientific production are important for the existence of chemometric societies and laboratories worldwide. Today, the world tends toward becoming more homogeneous with respect to chemometric activity, and will reach a corresponding normal distribution in about 70 years from now. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: chemometric history and etymology; languages; bibliometrics; webometrics; chemometric activity

1. INTRODUCTION

According to scholars of the history of chemistry [1–7], Christian Europe of the High Middle Ages was gradually adopting a new science of transformation of matter, together with its name, *alchemy*, from the Arabic civilization in the Middle East, Southern Italy, and the Iberian Peninsula. The Arabs called the science *al-kimiya* (the chemistry), the etymology of which is uncertain, and could be attributed to Medieval Chinese, Late Antiquity Hebrew, Late Greek, Ancient Egyptian, or even Coptic, from which the Arabic language could, directly or indirectly, take the word and prefix it with its definite article *al*. However, most scholars agree that the word *alchemy* originated either from the Egyptian *khem* or Greek *chemeia* or *chymea*. Medieval Latin absorbed *al-kimiya* as (*al*)*chymia* or (*al*)*chemia*, and further propagated it into all European languages (14th century English: *alchemy* [8]). The modern science was named *chemistry* in 18th century English [8]. The 20th century brought new terms beginning with combining forms **chemo-**, **chemi-**, or **chem-** [9], for example, *chemo*sphere, *chemo*therapy, *chemi*sorption, and *chem*informatics.

This is the time when there are already several scientific fields with composite names that end in *-metry* (like geo*metry*) and *-metrics* (like bio*metrics*), the suffixes derived from the Ancient Greek words for *measure* (*metron*), *measurement* (*metria*), or *to measure* (*metrein*). Very similar or related disciplines were named with equal first combining forms as, for example, *psycho-* in *psycho*metry and *psycho*metrics, and *econo-* in *econo*metry and *econo*metrics. This choice to name a scientific discipline with '*-ics*' (pronounced as [IKS]) or '*-y*' (pronounced as [I]) has become arbitrary in any European language. The choice seems not to be linguistically determined but a matter of widely accepted convention. In 1971, the Swedish word **kemometri** appeared for the first time, constructed by combining the forms **kemo-** for chemistry and **-metri** for measure, and its English equivalent **chemometrics** = **chemo-** + **-metrics**, both coined by Prof. S. Wold [10–12]. Since then, the new term has conquered its space in the names of chemical concepts, and is widely used by research groups, and in scientific meetings, journals, and disciplines (bold in Table I). The derivatives of

*Correspondence to: Prof. Dr M. M. C. Ferreira, Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas (UNICAMP), Campinas 13083-970, SP, Brazil.
E-mail: marcia@iqm.unicamp.br

†This paper was presented at the *10th International Conference on Chemometrics in Analytical Chemistry* (*CAC-2006*), Águas de Lindóia, SP, Brazil, 10–15 September 2006.

Copyright © 2007 John Wiley & Sons, Ltd.

Table I. Important etymological moments in the early history of chemometrics[a] (1971–1990)

Date	Event
1971	Prof. Svante Wold (Umeå University, Sweden) coins the term **kemometri**[b] in Swedish from **kemo** and **metri** and its English equivalent **chemometrics** (January) [10–14]
1972	Prof. S. Wold names his group Forskningsgruppen för **Kemometri** (Research Group for Chemometrics) [12] or **Kemometrigruppen** (Chemometrics Group), and publishes the first article using the term **kemometri** [15]
1973	The first article that has the term **chemometrics** in the name of a research group (Prof. S. Wold's group) is recorded in the SCI[c] database [16]
1974	Prof. S. Wold creates the definition: 'The art of extracting chemically relevant information from data produced in chemical experiments is given the name of '**chemometrics**' in analogy with biometrics, econometrics, etc.'[18]
1974	The International **Chemometrics** Society is founded by Professors S. Wold and Bruce R. Kowalski from the University of Washington, WA (June 10th) [10,12]
1975	The first article that has the term **chemometrics** in the title, and also the second article with **chemometrics** in the name of a research group, is published by Prof. B. R. Kowalski and recorded in the SCI database. The article defines **chemometrics** as a new chemical discipline [19]
1976	The first **Chemometrics** Newsletter is issued by Prof. B. R. Kowalski (January) [14]
1976	The first Symposium '**Chemometrics**: Theory and Applications' is organized by the Computers in Chemistry Division of the American Chemical Society in San Francisco, which results in the first proceedings volume (the first book) on **chemometrics**, published in 1977 [14,20]
1978	The International Conference on **Chemometrics** in Analytical Chemistry ('CAC Holland') is held in Petten, The Netherlands (September 15–17) [21]. This is called CAC-II even though it was the first to have **chemometrics** in its title
1979	The term **chemometric** is recorded for the first time in the WOS [22]
1980	The first review on **Chemometrics** appears in the Fundamental Reviews issue of Analytical Chemistry, as the successor to the reviews on Statistical and Mathematical Methods in Analytical Chemistry (April) [23]
1981	The term **chemometricians** is recorded for the first time in the WOS [24]
1984	**Chemometrics** is recorded for the first time in publications from Asia, as cited in the WOS: an article with **chemometrics** in the title [25], and another one with **chemometrics** in the address [26]
1986	**Chemometrics** and Intelligent Laboratory Systems is founded
1987	Journal of **Chemometrics** is founded
1987	The first Colloquium **Chemiometricum** Mediterraneum is held in Barcelona, Spain (November 9–11th) [27,28]
1988	The first professorship in **chemometrics** is seated at the Umeå University (Prof. S. Wold) [11]
1988	The First Scandinavian Symposium on **Chemometrics** is held in Lappeenranta, Finland (October 6–8th) [29]
1990	The early history of **chemometrics** is discussed by its founders [14,30]

[a] Composite expressions that contain the word **chemometrics** are underlined.
[b] More exactly, the term **kemometri** was published in Holmsund, 15 km from Umeå [13].
[c] SCI stands for Science Citation Index-Expanded, a part of the Web of Science (WOS) database of scientific publications [17].

chemometrics (noun plural in form but singular in meaning) include **chemometric** (adjective), **chemometrically** (adverb), and **chemometrician** (noun: a person trained in or dealing with chemometrics).

Sometimes a chemometrician or a scientist that employs chemometrics stops to think about the word **chemometrics**. Why is this discipline called this way? Is **chemometry** another name for it? How to say: **chemometric** or **chemometrical**? How is this word, **chemometrics**, written and pronounced in other languages? Why in Russian it is of the '*-ics*' type, and in French of the '*-y*' type, while in German and Czech it exists in both forms? Such and similar questions have initiated our small study on **chemometrics** in 16 languages in 2002 [31]. More extensive study was encouraged by **chemometrics** in 30 languages as recently published by K. Faber [32], **chemometrics/chemometry** discussion [33] that appeared at the beginning of this year on the e-mail List of the International Chemometric Society (ICS-L), and **chemometrics/chemometry** discussions or divided opinions in other languages [34–37]. These and related questions are, thus, the subject of the present work, where linguistic and bibliometric efforts are made to rationalize observed trends. Is the frequency of the word **chemometrics** in some language/country in a definite period of time some measure of chemometric activity and, consequently, of scientific and technological development of the country or countries in question? Does this frequency show a regular functional trend over time? Such questions are also considered. The present investigations can be useful for the chemometric community as an interesting historical, linguistic, and sociological viewpoint on **chemometrics**.

2. METHODS

Several database-mining searches in the Web of Science (WOS) database [17] and internet surfing using Google [38] and Yahoo [39] search engines were performed to obtain qualitative and quantitative data for bibliometric and linguistic studies of the word **chemometrics**.

2.1. Linguistic surfing

Among the search engines tested in this work, Google has shown to be the most capable to deal with Latin and diverse non-Latin fonts, whilst Yahoo was the second choice. They were used to find home pages of different kinds, preferentially official pages of research groups and individuals in science and higher education institutions worldwide, that possessed some chemometric material in their languages. These internet pages were considered as primary sources of information. The word **chemometrics** was searched for in various European and non-European languages (subsequently called: national languages) in its basic form

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

(**chemometrics**), and also in inflected forms whenever necessary. Official or widely accepted standard varieties of languages were taken into account, not local dialects. Special attention was paid to multicenter languages such as English, French, Swedish, and Chinese, among others. The results obtained were then refined by counting the frequency of pages for each of the forms for **chemometrics** in national languages. In case two or more forms existed in the same language, their relative frequencies were determined. The word **chemistry** in the same languages was also found by Google and Yahoo, in order to compare **chem-** in **chemometrics** and **chemistry**. Online dictionaries and extensive linguistic tools [40,41] were used to write down pronunciation for all encountered forms of **chemometrics** and **chemistry**, following the rules of the International Phonetics Association (IPA) [42] whenever possible. The forms in non-Latin writing systems were transliterated into the Latin system according to the IPA or other institution/authority that regulates a particular language.

2.2. Bibliometric search for international collaborations

Science Citation Index-Expanded (SCI) [17] of the WOS was used to determine the total number of scientific publications in the period 1945–2005 for countries in Europe, Asia, South Africa, and selected countries where English, Spanish, Portuguese, Russian, and Chinese are official languages in science and higher education (the older search for variable Tot, Table II). The number of publications from each country made in collaboration with other groups from different countries was determined and expressed as a fraction. The countries with collaborating groups and individual countries were defined based on linguistic, geographical, and historical bases, with the approximation that all countries were monolingual with either '*-ics*' (K) or '*-y*' (I,J) end types of **chemometrics**. The fractions for the end types were taken into account whenever they were greater than 4%. They were summed into %K and %IJ, respectively, to predict whether the country in question would, due to collaboration influences, create **chemometrics** of the K or I,J end type.

2.3. Other minings and surfings and chemometrics-development relationships

The SCI database was searched for publications with **chemometr*** as a topic (in title, keywords, and abstract; see descriptor Pub in Table II) for each year and additionally for each country, and also for publications with **chemometr*** as the address for each year. A new Google surfing was

Table II. Bibliometric, webometric, and development descriptors of 76 countries[a] worldwide

No.	Descriptor	Definition[b]	Method[c]
1	log(Tot)	Tot—Total No. scientific publications in 1945–2005 (22–25 May and 30 August 2006)	SCI
2	1stTot	1st publication date related to Tot (30 August 2006)	SCI
3	log(Pub)	Pub—No. publications with 'chemometr*' in topics (title, keywords, abstract) in 1975–2005 (23 March 2006)	SCI
4	1stpubl	1st publication date related to Pub (23 March 2006)	SCI
5	log(DNC + 1)	DNC—Annual No. publications for Pub in the last 5 years (23 March 2006)	SCI
6	log(JCpubs + 1)	JCpubs—No. publications in *J. Chemometr.* in 1987–2005 (31 August 2006)	SCI
7	log(CILSpubs + 1)	CILSpubs—No. publications in *Chemometr. Intell. Lab. Syst.* in 1986–2005 (31 August 2006)	SCI
8	log(Chempubs + 1)	Chempubs—No. publications in any of the two journals in 1986–2005 (31 August 2006)	SCI
9	log(WWW + 1)	WWW—No. hits for 'chemometrics' (22 March 2006)	Google
10	log(JCwww + 1)	JCwww—No. hits for 'Journal of Chemometrics' (31 August 2006)	Google
11	log(CILSwww + 1)	CILSwww—No. hits for 'Chemometrics and Intelligent Laboratory Systems' (31 August 2006)	Google
12	log(ChJwww + 1)	ChJwww—No. hits for any of the names of the two journals (1 September 2006)	Google
13	log(Ch1www + 1)	Ch1www—No. hits for 'chemometrics' and its 7 derivatives[d] (2 September 2006)	Google
14	log(Ch2www)	Ch2www—No. hits for 'chemistry' and its 6 derivatives[e] (2 September 2006)	Google
15	GDP	Gross Domestic Product per capita in US$ in 2004	Ref. [43–45]
16	RIRD	No. researchers in research and development per million people in 1999–2001	Ref. [46–50]
17	HDI	Human Development Index in 2002	Ref. [46–50]
18	Connec	Connectivity—Physical infrastructure for the information and communication technology in 2000	Ref. [51]
19	Access	Access—Wider determinants of access to the information and communication technology in 2000	Ref. [51]
20	Policy	Policy environment of the information and communication technology in 2001–2002	Ref. [51]
21	Diffusion	Diffusion of the information and communication technology in 2000	Ref. [51]
22	DAI	Digital Access Index, the overall ability to access and use information and communication technology in 2002	Ref. [52]

[a] The 76 countries for which all the descriptors were generated are those having at least one SCI (Science Citation Index-Expanded) publication with 'chemometr*' in topics (descriptor Pub).
[b] Dates of all searches are in parentheses.
[c] Method: SCI—bibliometric descriptors from SCI data minings, Google—webometric descriptors from Google internet surfings, Ref.—development indices from the literature with references.
[d] Derivatives of 'chemometrics': chemometric, chemometrical, chemometrically, chemometrician, chemometricians, chemometry, and chemometries.
[e] Derivatives of 'chemistry': chemical, chemicals, chemically, chemist, chemists, and chemistries.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

performed for the English word **chemometrics** for 250 countries and territories with recognized country domains (.xx) [53]. Additional seven bibliometric and six webometric descriptors related to chemometric activity were also generated from SCI and Google searches (Table II). Eight country development indices were used or estimated from the literature (Table II). The set of the 22 descriptors, from which many are in log form (Table II, including the second search for Tot), was organized into Europe data set (36 countries) and the world data set (76 countries), for those countries having non-zero values of Pub (Table II). Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) [54] using Pirouette [55] and Matlab [56] were carried out with autoscaled descriptors for both sets, in order to recognize the chemometric activity patterns in Europe and worldwide, and to show the existence of relationships between chemometrics (bibliometric and webometric descriptors) and development (country development indices).

3. RESULTS AND DISCUSSION

3.1. Chemometrics said in various languages: report

Table III summarizes the results of Google/Yahoo surfing with respect to the word **chemometrics** in various languages and writing systems, and the corresponding results for **chemistry**. **Chemometrics** was identified in 47 of the existing official languages in science and in higher education institutions worldwide, and in 1 regional language used in Indonesia (Sundanese, sun-1). This collection comprises 82 orthographic (written) forms for **chemometrics**, ranging from only one per language like in Swedish or Portuguese, to a maximum of six in English. Even more pronunciation forms (127) were encountered. **Chemometrics** was found written in 10 writing systems: Latin, Cyrillic, Greek, and Arabic alphabets, three Japanese scripts (Kanji, Hiragana, and Katakana), Chinese characters (Traditional, Simplified, and mixed), Korean alphabet (Hangul), and Thai alphabet. The term *language* was adopted in this work not only as a linguistic construct, but also as a historical, cultural, and political entity. The word **chemometrics** nicely illustrates that this approach was correct, as in case of Romanian–Moldovan and Russian–Ukrainian–Belarusan languages. **Chemometrics** was reconstructed for Lithuanian, Belarusan, and putative New Latin (Scientific), from available inflected forms and similarities with the closest languages. The 48 languages belong to distinct language families and, according to modern linguistic classification [57], are divided into the following language families and language isolates (language numbers in brackets).

1. Indo-European languages (35): Germanic (9), Romance (9), Slavic (12), Baltic (2), Iranian (1), Albanian (Tosk), and Greek sub-families or branches, as shown in Scheme 1.
2. Finno-Ugric languages (3): Estonian, Finnish, and Hungarian.
3. Altaic languages (1): Turkish.
4. Basque language.
5. Sino-Tibetan languages (1): Chinese Mandarin.
6. Thai-Kadai languages (1): Thai.
7. Austro-Asiatic languages (1): Vietnamese.
8. Austronesian languages (3): Indonesian, Malay, and Sundanese.
9. Korean language.
10. Japanese language.

The new interdisciplinary field of linguistics, genetics, and archeology [58–64] tends to unify language families 1–3 and language isolates 9 and 10 into a Eurasiatic superfamily [58,62,63], and families 6–8 into an Austric superfamily [60,62–64]. These linguistic relationships might point out the natural ease by which scientific terms can circulate among national languages. The 48 languages are spread over six continents: the Americas (English, French, Spanish, and Portuguese), Australia (English), Africa (Afrikaans in South Africa), Europe (38 languages), and Asia (12 languages including Russian and English). Geographic distribution of these languages together with other official languages in Europe and Asia is presented in Figures 1 and 2, respectively. All the languages are officially spoken *de jure* or *de facto* in these countries. Multilingual countries were drawn with putative (not exact) linguistic boundaries to denote the existence of distinct languages in the maps, using administrative and linguistic boundaries [40,57,65] as well as university linguistic preferences: Belarus, Belgium, Bosnia and Herzegovina, China, Cyprus, Finland, Indonesia, Kazakhstan, Kyrgyzstan, Luxembourg, Moldavia, Serbia, and Montenegro, Spain, Switzerland, and Ukraine. Other multilingual countries that are not native English speakers but use English extensively in science and public administration throughout entire national territory were colored only for English (Ireland and several South and South-East Asian countries, see Figures 1 and 2). Such a linguistic richness strongly indicates that new forms for **chemometrics** will appear soon. For the multilingual countries where **chemometrics** was not found in the minority languages, the data for the major official language were used in further analyses. Europe was almost entirely identified in terms of **chemometrics**. The Faroe Islands, an autonomous region of Denmark, were treated separately due to their linguistic and geographic distinction from Denmark.

3.2. Chemometrics in national languages: -ICS/-IKA versus -Y/IA?

From the 82 orthographic forms, 71 are written as only one word composed of **chemo-** and **-metrics** forms and, therefore, they can be used in comparative analysis. One of the recent subjects in the ICS-L was about the end forms of **chemometrics** in national languages, *that is*, if these ends were equivalent to English either '*-ics*' or '*-y*'. Table III confirms the existence of these two end types, even though the situation is rather complex. The word **chemometrics** *cannot* be characterized only by these two end forms. Most forms in the Romance and Slavic languages, in all Finno-Ugric and Baltic, in some Germanic languages, in Greek, Albanian, Basque, and even in Indonesian, prevent this simple classification. Besides the written forms, one should take into account the pronunciation forms of **chemometrics**. Figure 3 shows putative relationships among the orthographic (left) and

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Copyright © 2007 John Wiley & Sons, Ltd.

Table III. Linguistics of the words **chemometrics** and **chemistry** in various national languages

Smp.[a]	Language[b]	**chemometrics**[c]	Lat. transliterat.[d]	Pronunciation variants[e]	E[f]	Fq[g]	**chemistry**[h]	[chemistry][i]
afr-1	Afrikaans	chemometrie	chemometrie	[xə.moə.'mɛ.tri]	I	m	chemie, skeikunde	[xɛ.mi]
afr-2	Afrikaans	chemometrika	chemometrika	[xə.moə.'mɛ.tri.ka]	K	mi	chemie, skeikunde	[xɛ.mi]
alt-1	Albanian - Tosk	kemometria	kemometria	[kɛ.mɔ.mɛ.'tɾiaɛ]	J	u	kimia	['ki.mia]
bas-1	Basque	kimiometria	kimiometria	[kɪ.mɪ.o.'me.trɪa]	J	u	kimika	['kɪ.mɪ.ka]
blr-1	Belarusan[#]	хемометрія	khemometriya, chemometrija	[xɛ.mo.'mɛ.tri.ja]	J	h	хімія; chimija, khimiya	['xi.mi.ja]
blr-2	Belarusan[#]	хемометріка	khemometrika, chemometrika	[xɛ.mo.'mɛ.tri.ka]	K	h	хімія; chimija, khimiya	['xi.mi.ja]
bsn-1	Bosniac	hemometrija	hemometrija	[hɛ.mɔ.'mɛ.tri.ja]	J	u	hemija	['hɛ.mi.ja]
blg-1	Bulgarian	хемометрия	khemometrija, khemometriya	[xɛ.mɔ.'mɛ.tri.ja]	J	u	хімія; khemiya, khemija	['xi.mi.ja]
cvb-1	Catalan-Valencian-Balear	quimiometria	quimiometria	[ki.mju.mə.'trijə], [ki.mjɔ.me.'trijɑ]	J	u	química	['ki.mi.kə], ['ki.mi.kɑ]
cmn-1	Chinese - Mandarin	化學計量學	hua xue ji liang xue	[xuɑ.ɕuɤ.tɕi.liɑŋ.ɕuɤ], [xuɑ.ɕyɛ.tɕi.liɑŋ.ɕyɛ]	O	mi	huà xué	[xuɑ.ɕuɤ]
cmn-2	Chinese - Mandarin	化学计量学	hua xue ji liang xue	[xuɑ.ɕuɤ.tɕi.liɑŋ.ɕuɤ], [xuɑ.ɕyɛ.tɕi.liɑŋ.ɕyɛ]	O	mi	化学; hua xue	[xuɑ.ɕuɤ], [xuɑ.ɕyɛ]
cmn-3	Chinese - Mandarin	化學计量學	hua xue ji liang xue	[xuɑ.ɕuɤ.tɕi.liɑŋ.ɕuɤ], [xuɑ.ɕyɛ.tɕi.liɑŋ.ɕyɛ]	O	mi	化学; hua xue	[xuɑ.ɕuɤ], [xuɑ.ɕyɛ]
cmn-4	Chinese - Mandarin	hua xue ji liang xue	hua xue ji liang xue	[xuɑ.ɕuɤ.tɕi.liɑŋ.ɕuɤ], [xuɑ.ɕyɛ.tɕi.liɑŋ.ɕyɛ]	O	n	huà xüê	[xuɑ.ɕyɛ]
crt-1	Croatian	kemometrija	kemometrija	[kɛ̝.mo̝.'mɛ̝.tri.ja]	J	ah	kemija	['kɛ̝.mi.ja]
crt-2	Croatian	kemometrika	kemometrika	[kɛ̝.mo̝.'mɛ̝.tri.ka]	K	bh	kemija	['kɛ̝.mi.ja]
cze-1	Czech	chemometrie	chemometrie	[xɛ.mo.'mʲɛ.trɟɛ]	J	m	chemie	['xɛ.miɛ]
cze-2	Czech	chemometrika	chemometrika	[xɛ.mo.'mʲɛ.trɟ.ka]	K	mi	chemie	['xɛ.miɛ]
dan-1	Danish	kemometri	kemometri	[kʰɛ.mɔ.mɛ.'tʁi]	I	au	kemi	[kʰɛ.'mi]
dan-2	Danish	kemometrik	kemometrik	[kʰɛ.mɔ.mɛ.'tʁiɡ̊]	K	mt	kemi	[kʰɛ.'mi]
dut-1	Dutch	chemometrie	chemometrie	[xə.mɔ.'mɛ.tʁɪə]	J	u	scheikunde, chemie	['xɛ.mɪə]
eng-1	English	chemometrics	chemometrics	[ˌkɛ.mə.'mɛ.tɹɪks], [ˌki ː .mə.'me.tɹɪks], [ˌkɛ.mo.'mɛ.tɹɪks], [ˌki.mo.'mɛ.tɹɪks],	K	au	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]

Continues

J. Chemometrics 2006; **20**: 247–272
DOI: 10.1002/cem

Table III. (Continued)

Smp.[a]	Language[b]	**chemometrics**[c]	Lat. transliterat.[d]	Pronunciation variants[e]	E[f]	Fq[g]	**chemistry**[h]	[chemistry][i]
				[ˌke.məʉ.ˈme.tɹɪks], [ˌkiː.məʉ.ˈme.tɹɪks]				
eng-2	English	chemometry	chemometry	[ˌkə.ˈmɒ.mɪ.tɹɪ], [ˌkɪ.ˈmɒ.mɪ.tɹɪ], [ˌkə.ˈmɑ.mə.tɹɪ], [ˌkɪ.ˈmɑ.mə.tɹɪ], [ˌkə.ˈmɔ.mɪ.tɹɪ], [ˌkɪ.ˈmɔ.mɪ.tɹɪ]	I	mt, n*	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]
eng-3	English	chemiometrics	chemiometrics	[ˌkɛ.mɪə.ˈmɛ.tɹɪks], [ˌkiː.mɪə.ˈmɛ.tɹɪks], [ˌkɛ.mɪɹ.ˈmɛ.tɹɪks], [ˌki.mɪɹ.ˈmɛ.tɹɪks], [ˌke.mɪə.ˈme.tɹɪks], [ˌkiː.mɪə.ˈme.tɹɪks]	K	n	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]
eng-4	English	chemiometry	chemiometry	[ˌkɛ.mɪ.ˈɒ.mɪ.tɹɪ], [ˌkɪ.mɪ.ˈɒ.mɪ.tɹɪ], [ˌkɛ.mɪ.ˈɑ.mɪ.tɹɪ], [ˌkɪ.mɪ.ˈɑ.mɪ.tɹɪ], [ˌke.mɪ.ˈɔ.mɪ.tɹɪ], [ˌkɪ.mɪ.ˈɔ.mɪ.tɹɪ]	I	n	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]
eng-5	English	chemimetrics	chemimetrics	[ˌkɛ.mɪ.ˈmɛ.tɹɪks], [ˌkɪ.mɪ.ˈmɛ.tɹɪks], [ˌkɛ.mɪ.ˈmɛ.tɹɪks], [ˌkɪ.mɪɹ.ˈmɛ.tɹɪks], [ˌke.mɪ.ˈme.tɹɪks], [ˌkɪ.mɪ.ˈme.tɹɪks]	K	n	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]
eng-6	English	chemimetry	chemimetry	[ˌkə.ˈmiː.mɪ.tɹɪ], [ˌkɪ.ˈmiː.mɪ.tɹɪ], [ˌkə.ˈmi.mɪ.tɹɪ], [ˌkɪ.ˈmi.mɪ.tɹɪ], [ˌke.ˈmiː.mɪ.tɹɪ], [ˌkɪ.ˈmiː.mɪ.tɹɪ]	I	n	chemistry	[ˌkɛ.mə.stɹɪ], [ˌkɛ.mɪ.stɹɪ], [ˌke.mə.stɹɪ], [ˌke.mɪ.stɹɪ]
est-1	Estonian	kemomeetria	kemomeetria	[ke.mo.ˈme.tria]	J	u	keemia	[ˈke.mia]
far-1	Faroese	kemometri	kemometri	[çɛ.mɔ.ˈmeː.trɪ]	I	u	evnafrøði, kemi	[çɛ.mɪ]
fin-1	Finnish	kemometria	kemometria	[ke.mo.ˈme.triɑ]	J	u	kemia	[ˈke.mia]
fre-1	French	chimiométrie	chimiométrie	[ʃi.mi̯.ɔ̃.ˈme.tʁi], [ʃi.mi̯.õ.ˈme.tʀɪ]	I	am	chimie	[ʃi.mi]
fre-2	French	chimiometrie	chimiometrie	[ʃi.mi̯.ɔ̃.mɛ.ˈtʁi]	I	mt	chimie	[ʃi.mi]
fre-3	French	chemiométrie	chemiométrie	[ʃɛ.mi̯.ɔ̃.ˈme.tʁi]	I	n	chimie	[ʃi.mi]
fre-4	French	chémiométrie	chémiométrie	[ˌʃe.mi̯.ɔ̃.ˈme.tʁi]	I	n	chimie	[ʃi.mi]
gal-1	Galician	quimiometría	quimiometría	[ki.mjo.mɛ.ˈtrija]	J	u	química	[ˈki.mi.ka]
ger-1	German	Chemometrie	Chemometrie	[ɕɛ.mo.me.ˈtriː], [kɛ.mo.me.ˈtriː], [ʃɛ.mo.me.ˈtriː]	I	am	Chemie, Scheikunde, Scheidekunst	[ɕɛ.ˈmiː], [kɛ.ˈmiː], [ʃɛ.ˈmiː]

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Copyright © 2007 John Wiley & Sons, Ltd.

ger-2	German	Chemometrik	Chemometrik	[ɕɛ.mo.me.'tri ː k], [kɛ.mo.me.'tri ː k], [ʃɛ.mo.me.'tri ː k]	K	mt	Chemie, Scheikunde, Scheidekunst	[ɕɛ.'mi ː], [kɛ.'mi ː], [ʃɛ.'mi ː]
grk-1	Greek	χημειομετρία	chimeiometria, khimiometria	[çi.mi.o̝.me̝.'tria]	J	u	χημεία; chimeia, khimeia	[çi.'mia]
hng-1	Hungarian	kemometria	kemometria	[kɛ.mo.'mɛ.t̪r̪iɒ]	J	u	kémia, vegytan	['ke ː .miɒ]
icl-1	Icelandic	chemometrics	chemometrics	[xɛ.mɔ.mɛ.'tri ː ks]	K	u	efnafræði, kemi	[ke.mi]
ind-1	Indonesian	kemometri	kemometri	[ˌke.mo.me.'tri]	I	h	kimia	['ki.mia]
ind-2	Indonesian	kemometrik	kemometrik	[ˌke.mo.'me.trik]	K	bh	kimia	['ki.mia]
ind-3	Indonesian	kemometrika	kemometrika	[ˌke.mo.'me.tri.ka]	K	mt	kimia	['ki.mia]
itl-1	Italian	chemiometria	chemiometria	[kɛ.mijo ː .me.'trija]	J	am	chimica	['ki.mi.ka]
itl-2	Italian	chimiometria	chimiometria	[ki.mijo ː .me.'trija]	J	n	chimica	['ki.mi.ka]
jpn-1	Japanese	ケモメトリックス	kemometoritsukusu	[ke̝.mo̝.'me̝.to̝.ɻi̲ 'tsü̞.kü̞.sü̞]	O	am	ケミストリー; kemisutori	[ke̝.mi.'sü̞.to̝.ɻi̲]
jpn-2	Japanese	計量化学	keiryou kagaku	['ke̞i̞..ɻi̲o̝.ü̞ 'kä.gä.kü̞]	O	mt	化学; kagaku	['kä.gä.kü̞]
jpn-3	Japanese	けいりょうかがく	keiryou kagaku	['ke̞i̞..ɻi̲o̝.ü̞ 'kä.gä.kü̞]	O	n	かがく; kagaku	['kä.gä.kü̞]
jpn-4	Japanese	kemometori	kemometori	[ke̝.mo̝.'me̝.to̝.ɻi̲]	I	n	kemisutori	[ke̝.mi.'sü̞.to̝.ɻi̲]
kor-1	Korean	계량분석화학	yae ryang pun seok hao hak, yay ryang pun sek hao hak	[jei.ɾjaŋ.pun̪.sɔk.hɑo.hɑk]	O	u	화학; hao hak	[hɑo.hɑk]
lat-1	Latin#	chemometria	chemometria	[kʰe.mo.'me ː .tria]	J	m	chemica, chemia, chymia	['kʰe.mi.ka], ['kʰe.mia], ['kʰi.mia]
lat-2	Latin#	chimiometria	chimiometria	[kʰi.mjo.'me ː .tria]	J	mi	chemica, chemia, chymia	['kʰe.mi.ka], ['kʰe.mia], ['kʰi.mia]
lav-1	Latvian	hemometrija	hemometrija	[xɛ.muo.'mɛ.tri.jɑ]	J	h	ķīmija	['ci ː .mi.jɑ]
lav-2	Latvian	kemometrija	kemometrija	[kɛ.muo.'mɛ.tri.jɑ]	J	bh	ķīmija	['ci ː .mi.jɑ]
lav-3	Latvian	kemometrika	kemometrika	[kɛ.muo.'mɛ.tri.kɑ]	K	mi	ķīmija	['ci ː .mi.jɑ]
lit-1	Lithuanian#	chemometrija	chemometrija	[xɛ.mo.'mɛ.tri.jɑ]	J	u	chemija	['xɛ.mi.jɑ]
mcd-1	Macedonian	хемометрија	hemometrija	[hɛ.mɔ.'mɛ.tri.ja]	J	u	хемија; hemija	['hɛ.mi.ja]
mal-1	Malay	kimometrik	kimometrik	[ˌki.mo.'me.trik]	K	m	kimia	['ki.mia]
mal-2	Malay	kemometrik	kemometrik	[ˌke.mo.'me.trik]	K	mi	kimia	['ki.mia]
mld-1	Moldovan	hemometrica	hemometrica	[he.mo.'me.tri.ka]	K	u	chimie	['ki.mʲe]
nrw-1	Norwegian**	kjemometri	kjemometri	[ɕɛ.mɔ.mə.'tɾi]	I	u	kjemi	[ɕɛ.mi]
per-1	Persian	شیمی‌سنجی	shimiyā' sinūgiyā', šimiyā' sinūžiyā'	[ʃi ː .mi.'jɒ ː si ː .nu ː .ʤi ː .'jɒ ː]	O	u	شیمی; shimiyā, šimiyā'	[ʃi ː .mi.'jɒ ː]

Continues

J. Chemometrics 2006; **20**: 247–272
DOI: 10.1002/cem

Table III. (Continued)

Smp.[a]	Language[b]	**chemometrics**[c]	Lat. transliterat.[d]	Pronunciation variants[e]	E[f]	Fq[g]	**chemistry**[h]	[chemistry][i]
pol-1	Polish	chemometria	chemometria	[xɛ.mɔ.'mɛ.trja]	J	u	chemia	['xɛ.mja]
prt-1	Portuguese	quimiometria	quimiometria	[kĩ.mjõ.me.'tɾia], [kĩ.mjũ.mɨ.'tɾiɐ]	J	u	química	[kĩ.mi.ka], [kĩ.mi.kɐ]
rmn-1	Romanian	chemometrie	chemometrie	[kɛ.mɔ.mɛ.'tɾie]	J	am	chimie	['ki.mje]
rmn-2	Romanian	chemometria	chemometria	[kɛ.mɔ.mɛ.'tɾia]	J	mt	chimic	['ki.mje]
rus-1	Russian	хемометрика	khemometrika	[xʲɛ.mo.'mʲɛ.tri.ka]	K	am	химия; khimiya, khimija	['xʲi.mi.ja]
rus-2	Russian	хемометрия	khemometrija, khemometriya	[xʲɛ.mo.'mʲɛ.tri.ja]	J	mt	химия; khimiya, khimija	['xʲi.mi.ja]
srb-1	Serbian	хемометриј**а**, hemometrija	hemometrija	[hɛ.mɔ.'mɛ.tri.ja]	J	m	хемија, hemija	['hɛ.mi.ja]
srb-2	Serbian	хемометрик**а**, hemometrika	hemometrika	[hɛ.mɔ.'mɛ.tri.ka]	K	mi	хемија, hemija	['hɛ.mi.ja]
slk-1	Slovak	chemometria	chemometria	[xɛ.mɔ.'mʲɛ.trja]	J	u	chémia	['xɛ.mja]
slv-1	Slovenian	kemometrija	kemometrija	[ke.mɔ.me.'tɾi.jɑ]	J	au	kemija	['ke.mi.jɑ]
slv-2	Slovenian	kemometrika	kemometrika	[ke.mɔ.'me.tɾi.kɑ]	K	n	kemija	['ke.mi.jɑ]
spn-1	Spanish	quimiometría	quimiometría	[ki.mjo.me.'ʈɾija], [ki.mjo.me.'ʈɾɪjɑ]	J	am	química	['ki.mi.ka], ['ki.mi.kɑ]
spn-2	Spanish	quimiometria	quimiometria	[ki.mjo.'me.ʈɾija], [ki.mjo.'me.ʈɾɪjɑ]	J	mt	química	['ki.mi.ka], ['ki.mi.kɑ]
sun-1	Sundanese	kémométrik	kémométrik	[ke.mõ.mẽ.'trik]	K	u	kimia	['ki.mĩã]
swe-1	Swedish	kemometri	kemometri	[ɕe.mʊ.me.'tri ː], [ɕe.mʊ.me.'tʁi ː], [tɕe.mʊ.me.'tʁi ː]	I	u	kemi	[ɕe.mi]
tha-1	Thai	**การตรวจวัดทางเคมี**	kart rw cw_4kh thang khemi	[ka ː rt.rw.'tɕwkʰ.tʰa ː ŋ.kʰe ː .mi ː]	O	m	**วิชาเคม**; wicha khem	[wi.tɕʰa ː .kʰe ː m]
tha-2	Thai	**เคโมเมทริกส์**	khemomethriks	[kʰe ː .mo ː .'me ː .tʰriks]	K	mi	**วิชาเคม**; wicha khem	[wi.tɕʰa ː .kʰe ː m]
trk-1	Turkish	kemometri	kemometri	[ke.mø.me.'tri]	I	u	kimya	[ki.'mja]
ukr-1	Ukrainian	хемометрія	khemometriya, khemometrija	[xʲɛ.mɔ.'mʲɛ.tri.ja]	J	m	хімія, khimiya, khimija	['xʲi.mi.ja]
ukr-2	Ukrainian	хемометрика	khemometrika	[xʲɛ.mɔ.'mʲɛ.tri.ka]	K	mi	хімія, khimiya, khimija	['xʲi.mi.ja]
vie-1	Vietnamese	thước đo hoá học	thước đo hoá học	[tʰɯ.'ə ː k.ɗɔ.'hɔɐ ː .'hok]	O	u	hoá học	['hɔɐ ː .'hok]

[a]Smp.—sample: a written form of the word **chemometrics** in different languages, as found on the internet. It is given by the language code adopted here and ordinary numbers for this language (as arranged in decreasing frequency order).
[b]Language—language name as recognized in linguistic, historical, cultural and political terms.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

pronunciation (right) ends. The 12 orthographic end form types can be classified as K, I, or J because after [TRI] (in Japanese: [TORI]), they can contain a consonant [K] (K), have no additional sound (I), or possess two vowels (J). From here on, a general representation of similar sounds will be marked with a capital letter and particular forms with a small letter, always in square brackets to denote pronunciation. There are much more pronunciation end form types (46) when short and long (**:**) vowels, and unstressed and stressed (ʹ) syllables are taken into consideration. It is visible that the I forms (equivalent to English '*-y*' or [I]) are between the K and J forms.

The K forms make two sub-groups. In K_m, the end forms have the voiceless consonant [k] or its voiced equivalent [G], sometimes followed by another voiceless consonant [s]. These end forms are monosyllabic. The other sub-group, K_b, contains disyllabic end forms, due to formation of additional syllable from [k] and a final [A]. This syllable is never stressed. Thus, K_b forms, unlike K_m forms, could not be considered exactly as equivalents to the English '*-ics*' ([IKS]). In all J forms, due to the existence of the combination of [I] with another vowel (not always diphthongs), the semi-vowel [J] (as the first sound [j] in English word 'yes') is frequently present. The semi-vowel [j] forms an additional syllable with [A] in several Slavic and Baltic languages. Figure 3 illustrates the continuous transition between all I and J forms, and the existence of a smaller gap (no intermediates) between K_m and K_b and larger one between K_b and J. The qualifier Fq in Table III classifies the 71 orthographic forms of **chemometrics**, taking into account the classes from Figure 3.

With respect to the number of syllables in **chemometrics**, most of the 71 forms are four-syllable words like the English **che-mo-me-trics** ([kE.mo.mE.tRıks]). Exceptions are five-syllable forms from the K_b sub-group, from the J sub-group with a [jA] syllable, the Japanese (jpn-4, Table III), and some forms that possess an inserted syllabic vowel between the central syllables [mO] and [mE]: Basque (bas-1), French (fre-1 to fre-4), and Greek (grk-1) forms.

Can scientists from different parts of the world understand each other's **chemometrics** (the 71 forms)? It depends on how much the basic structure of the forms is conserved. The orthographic forms can be written as XEmOmetr*, where X represents the main variation (ch, k, kh, q, h, kj), the two 'm' letters are preserved, O has some moderate variation (o, io, eio, i), 'e' is preserved (e or é), 'tr' is also preserved (tr, thr, tor), and '*' is the end that has been already discussed (Figure 3 left). The pronunciation forms show more variations and can be generalized as [XEmOmEtRI*], with the largest variations in [O] and the ends [*] (Figure 3 right), significant in [I], moderate in [X] and two [E] sounds, and smallest in [tR]. Concluding, the scientists could understand each other's **chemometrics** in written form, and also when pronouncing them clearly and not too fast.

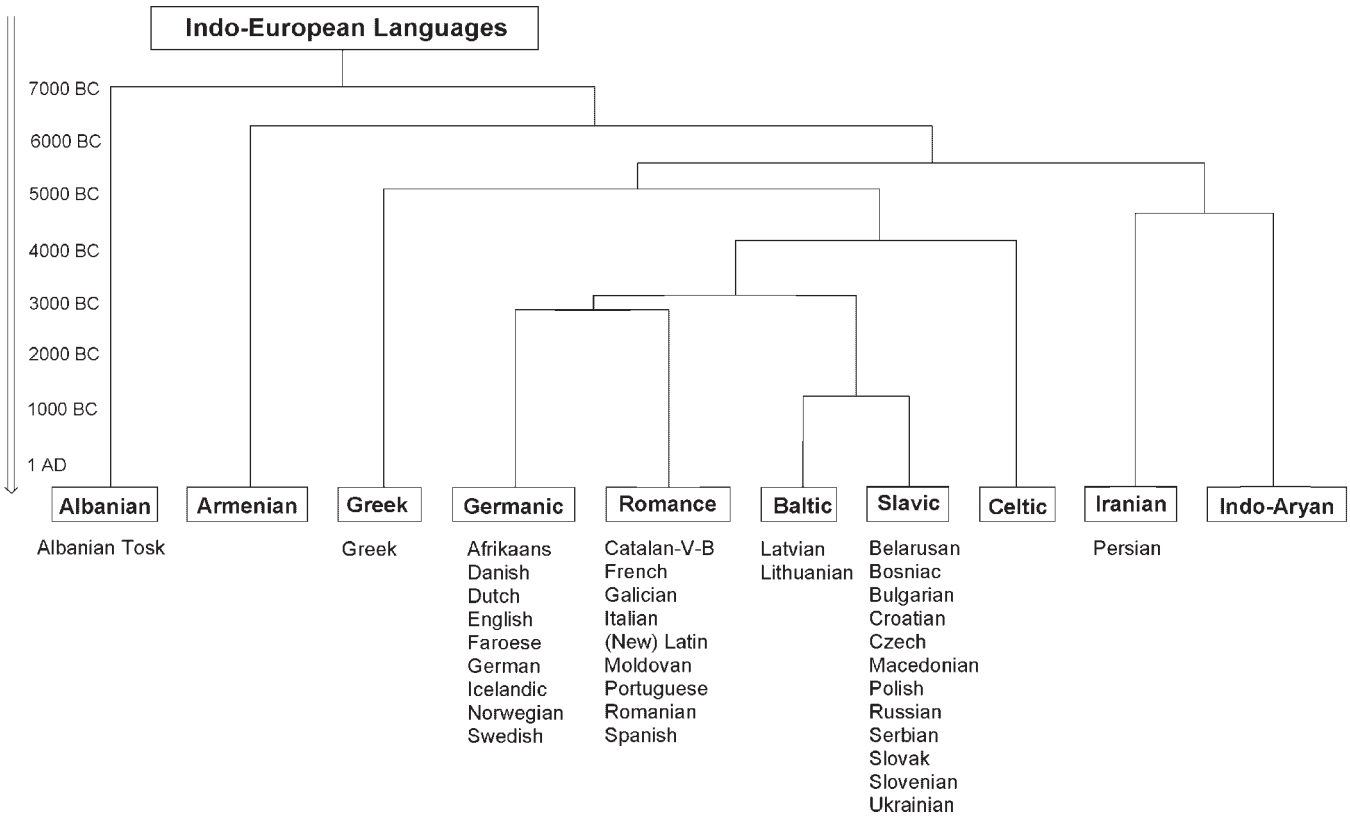
3.3. Chemometrics in Asia: linguistic situation reveals historical and modern ties

Since the word **chemometrics** had been coined in English and Swedish in 1971, countries other than Sweden and English-speaking countries had to name the new discipline in their language(s). To adopt a foreign form or create one's

[c]**Chemometrics**—a written form of the word **chemometrics** in a particular language.
[d]Lat. transliterat.—the most common way(s) of Latin transliteration for forms written in non-Latin writing systems; for those written in Latin-based alphabets, the forms are equal to the original forms.
[e]Pronunciation variants—the most frequently or most important official pronunciation variants in standard(s) of a particular language. The following languages have these standards as arranged in decreasing order of frequencies of the pronunciation forms of **chemometrics**: (a) Catalan–Valencian–Balear: 1—Eastern-Central Catalan, 2—Valencian and Balearic; (b) Chinese Mandarin: 1—Mandarin from the PR China, 2—Mandarin from Taiwan; (c) English: 1—two British variants (Received Pronunciation), 2—two General American variants, 3—two Australian English variants; (d) French: 1—standard French from France, Belgium and Switzerland, 2—Québec French; (e) German: 1, 2, 3—three common pronunciation variants with unknown relative frequencies; (f) Portuguese: 1—Brazilian Portuguese, 2—European Portuguese; (g) Spanish: 1—standard Spanish, 2—Andalusian Spanish; (h) Swedish: 1—Central Swedish standard, 2—South Swedish standard, 3—Finland–Swedish standard.
[f]E—the pronunciation end types: I (the end with long or short [I]-type sound), J (end with semi-wovel [j] or vowel combinations as in [iə], [ia], [jə], [ja], etc.), K (end with [k]-type sound as [ks], [ka], or [k]), and O (other end forms in morphologically different words—in Asian languages).
[g]Fq—relative frequency of the samples: u—unique (100%), au—almost unique (>99%), am—absolute majority (90–99%), m—major (60–90%), ah—above half (53–60%), h—around half (47–53%), bh—below half (40–47%), mi—minor (10–40%), mt—minority (0.5–10%), and n—negligible (<0.5%).
[h]**Chemistry**—written forms of the word **chemistry**, in decreasing order of frequency for a particular language.
[i][chemistry]—pronunciation variant for the word **chemistry** that is morphologically closest to form for **chemometrics**, according to the same standard language rules.
[*]Fq qualifier values for UK and USA only.
[#]Reconstructions from inflected forms (Latin and Latvian) or forms from the second official language (Russian in Belarus).
[**]Bokmål and Nynorsk are not distinguished here.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem



Scheme 1. Indo-European family of languages with its living sub-families (branches) typed in bold. Particular languages are named below the corresponding branches when **chemometrics** is found in these languages (Table II). Catalan-V-B stands for Catalan-Valencian-Balear. More precisely, Romance is a sub-branch which, together with Latin, forms the Italic branch. The dendogram representation of similarities among the languages was adapted from existing lexicostatistical studies [58,59]. It also represents historical branching of the Indo-European family during 9 millennia [58], as shown by the time scale left to the dendogram.

own? To use one or more expressions? According to Ojasso [66], 'although a nation can form part of a group with typical features, it never relinquishes its individuality' in scientific production. However, some general mechanisms of foreign influence in a particular country in generating scientific terms may be pointed out [66].

1. Linguistic and genetic relationships: countries that speak the same or similar languages (i.e., from the same linguistic branch), and probably have common ancestors [66,68];
2. Geographical relationships: neighbors or other close countries;
3. Traditional historical, cultural, economic, scientific, and political relationships [66];
4. Scientific, technological, and economic development of the country in question;
5. International scientific collaboration with generally increasing tendencies [67–70]: publications, projects, formation of scientists, etc., jointly with other countries;
6. Presence of other first language speakers: these speakers tend to use foreign terminology.

The root for chemistry, **chem-** became the heritage of the Roman civilization and its successors (Occidental, Orthodox, and Muslim civilizations). **Chem-** entered into other civilizations in the 19–20th centuries. However, some Asian civilizations maintained their original peculiarity in science and, consequently, in naming chemistry and chemometrics. For example, former countries and territories dependent on the UK and USA in Asia (see Figure 2) accepted English in science and higher education, and also the words **chemistry** and **chemometrics**. In a similar way, some former USSR members in Asia, with a large number of native Russian speakers, use Russian in science, and the expressions **khimiya** and **khemometrika**. The Turkish **kemi/kemometri** and the Austronesian forms for **chemometrics** and **chemistry** came obviously from European languages. South and South-East Asia were dependent on some European powers and the USA for relatively long times and through this have been affected linguistically. The leadership of North America, Western Europe, and Japan in science and technology makes these countries as the most frequent partners in international collaborations with this part of Asia [69–71]. The Iranian **shimiya** comes from Arabic **al-kimiya**, but the expression for **chemometrics** consists of two words (per-1, Table III) meaning 'data analysis chemistry'.

The Chinese word for chemistry is **hua xue**, literally meaning **transformation study**. The five short words **hua xue ji liang xue** may be translated as **quantification methods in chemistry**. Korean, Japanese, and Vietnamese, although

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Figure 1. Official science and higher education languages in Europe and their linguistic classification.

linguistically distant from Chinese and each other, were under Chinese cultural and linguistic influence for a long time (including the usage of Chinese characters that only ceased in Vietnam some two decades ago). This is visible from the form for **chemistry** that is present in the expressions for **chemometrics**: Japanese: **kagaku**, Korean: **hao hak**, and Vietnamese: **hóa học**. Some Asian languages whose basic writing systems are non-Latin, adopted one-word equivalents for **chemometrics** that are well-recognizable (Figure 3): two Japanese (jpn-1 and jpn-4) and two Thai forms (tha-1 and tha-2) use **khem-** for **chemistry** and/or **chemometrics**.

Comparing **chemometrics** and **chemistry** in Asian languages, one can notice that a Malay form (mal-2), the Sundanese (sun-1), three Indonesian forms, and the Turkish (trk-1) were not created from **chemistry** in these languages. The existence of more than one transliterated/Latin form for **chemometrics** in Japanese, Indonesian, and Thai indicates foreign influences. It is known [72,73] that the size of scientific production and some extra-science factors such as history, geography, politics, language, and dependence of one country on another determine who collaborates with whom in the international scientific community. Table IV represents international collaboration levels with countries that might affect the creation of K (%K) or I,J (%IJ) orthographic forms for **chemometrics** in a particular country. It is seen that the predictions (the last columns in Table IV) for the K/I,J choice are negative, *that is*, badly predicted for all countries with the total of 100 000 or more scientific publications, except for some English-speaking countries due to traditionally intense collaborations among them (right top corner of the correlogram in Figure 4). Asian countries speaking national languages and not English in science also do not have good predictions, with the exception of Thailand (*-iks* or K form was predicted), Malaysia (*-ik* or K form was

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

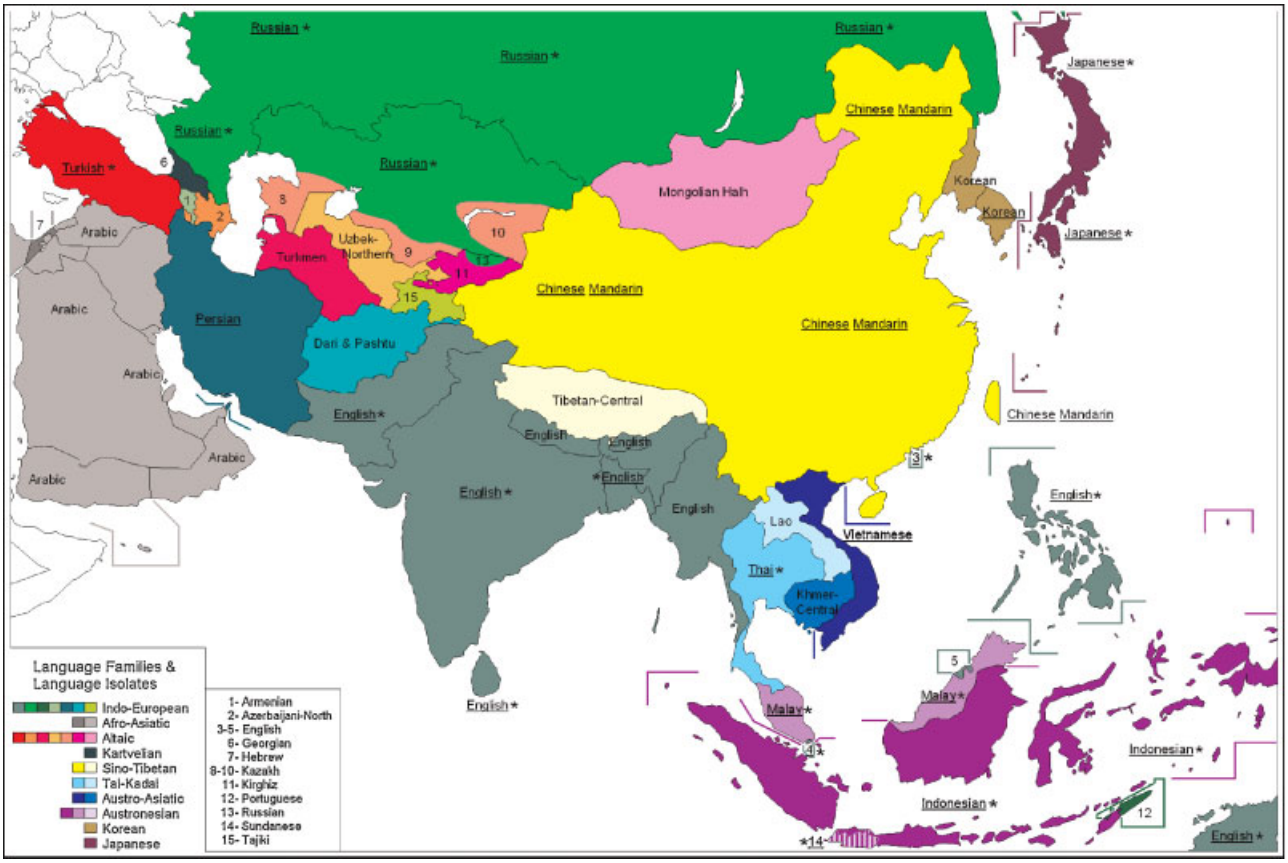


Figure 2. Official science and higher education languages in Asia and their linguistic classification. Underlined names indicate those languages for which the word **chemometrics** was found in online documents. Names with asterisk (*) represent languages in which at least one form for **chemometrics** was found to consist of **chemo-** and **-metrics** only.

predicted correctly), and Indonesia (all *-ik* or K forms were also predicted correctly). In general, English-speaking countries are the major contributors to K forms, and Western Europe and Japan to the I,J forms. Obviously, countries with scientific production below 100 000 publications show better prediction, *that is*, they are probably more influenced by countries with which they have significant scientific collaboration (Figure 4).

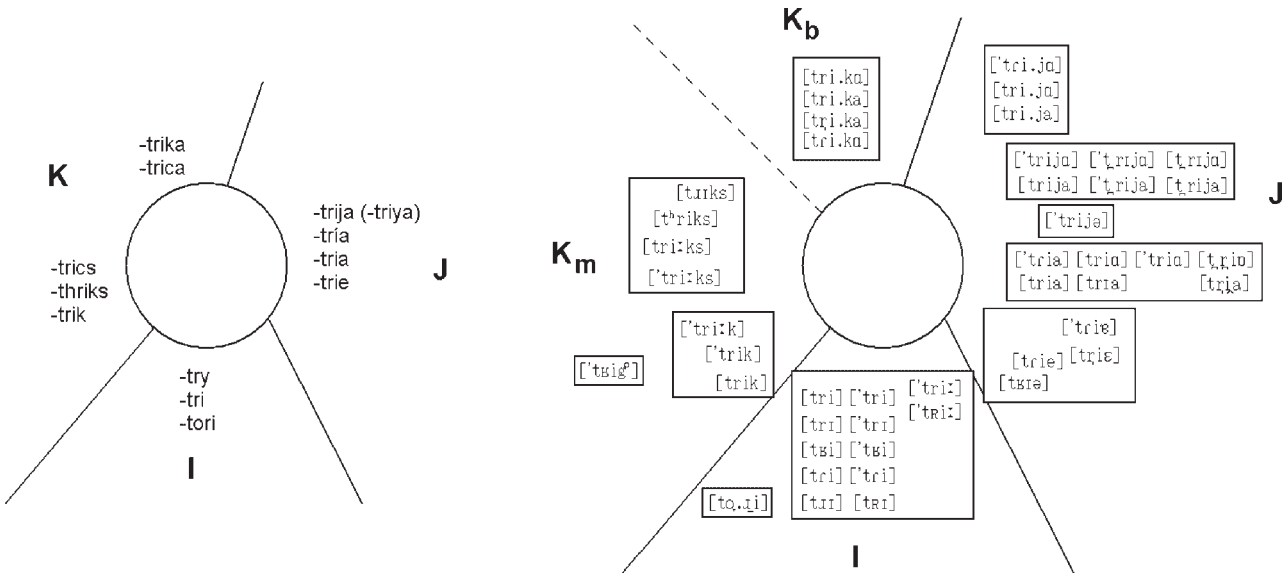


Figure 3. Putative classification of orthographic (left) and pronunciation (right) end forms of the word **chemometrics** in national languages.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Table IV. Possible effect of international scientific collaborations on the K/I,J choice for the end forms for **chemometrics** in national languages and countries

Country/language[a]	Tot[b]	%K[c]	%IJ[c]	Predicted[d]	+/−	Country/language[a]	Tot[b]	%K[c]	%IJ[c]	Predicted[d]	+/−
USA/eng	≈6000[#]	6	6	-ics	+	Turkey/trk	104	11	0	-ik	–
UK/eng	2064	15	4	-ics/(-y)	+–	Ireland/eng	104	51	9	-ics/(-y)	+–
Japan/jpn	1703	8	0	-i	–	Greece/grk	99	18	15	-ika/-ia	–
France/fre	1280	12	9	-iques/-ie	–	Argentina/spn	95	12	13	-ica/-ia	–
Germany/ger	1143	15	7	-ik/(-e)	–	Czech Rep./cze	59	20	31	-ika/-ie	+
Canada/eng	1009	17	0	-ics	+	Portugal/prt	54	20	29	-ica/-ia	+–
Italy/itl	712	15	12	-ica/-ia	–	Chile/spn	49	19	21	-ica/(-ia)	–
Australia/eng	525	25	0	-ics	+	Bulgaria/blg	47	8	15	(-ika)/-ija	+–
India/eng	512	6	0	-ics	+	Romania/rmn	43	10	18	(-ica)/-ie,-ia	+–
Netherlands/dut	476	16	12	-ik/-ie	–	Iran/per	28	17	0	-ika	–
China/cmn	469	14	0	-ik	–	Thailand/tha	26	31	15	-ks/-i	+–
Spain/bas + cvb + gal + +spn	435	12	13	-ica/-ia	–	Slovakia/slk	25	20	41	(-ka)/-ie	+
Sweden/swe	398	15	15	-ik/-i	–	Slovenia/slv	18	19	32	(-ka)/-ija	+–
Russia/rus	389	9	13	-ika/-ija	–	Malaysia/mal	17	28	5	-ik/(-ia)	+
Switzerland/ger	367	19	22	-ik/-ie	–	Croatia/crt	16	14	18	-ika/-ija	+
Switzerland/fre	367	19	22	-iques/-ie	–	Estonia/est	9	16	39	(-ika)/-ia	+–
Switzerland/itl	367	19	22	-ica/-ia	–	Lithuania/lit	9	15	26	(-ka)/-ija	+
Belgium/dut	255	15	22	-ik/-ie	–	Indonesia/ind	8	43	38	-ik/-i	+
Belgium/fre	255	15	12	-iques/-ie	–	Indonesia/sun	8	43	38	-ik/-i	+–
Poland/pol	240	11	14	-ika/-ia	–	Latvia/lav	7	16	24	(-ika)/-ija	+
Denmark/dan	206	20	20	-ik/-i	–	Kazakhstan/rus	6	17	0	-ika	+
Brazil/prt	200	14	7	-ica/-ia	–	Iceland/icl	6	31	38	-ics/-i	–
South Korea/kor	199	17	5	-ik	–	Vietnam/vie	6	25	40	(-ik)/-i	–
Austria/ger	173	13	12	-ik/-ie	–	Serbia&Mon./srb	4	9	12	-ika/-ija	+
Finland/fin	166	15	18	-ika/-ia	–	Belarus/blr + rus	3	17	20	-ika/-ija	+
Taiwan/cmn	159	13	0	-ik	–	Luxembourg/fre	2	14	50	(-iques)/-ie	+
Mexico/spn	150	47	7	-ica/(-ia)	–	Luxembourg/ger	2	14	50	(-ik)/-ie	+
Norway/nrw	124	13	21	(-ik)/-i	–	Moldavia/mld	2	7	0	-ica	+
Hungary/hng	120	14	14	-ika/-ia	–	Macedonia/mcd	2	18	13	-ika/-ja	–
South Africa/afr	116	14	0	-ika	–	Albania/alt	1	11	52	(-ka)/-ia	+
New Zealand/eng	108	21	0	-ics	+	Bosnia&Her./bsn	1	31	41	(-ika)/-ija	+–
Ukraine/ukr	106	9	5	-ika/(-ija)	–	Faroe Islands/far	<1	42	81	(-ik)/-i	+

*Collaborating groups of countries or individual countries: (i) with predominant K forms for **chemometrics**: English-speaking countries (USA, UK, Canada, Australia, New Zealand, South Africa, and Ireland), Asian countries with official English in science (India, Sri Lanka, Pakistan, Myanmar, Bangladesh, Singapore, Hong Kong, and The Philippines), and East Slavic countries (Russia, Ukraine, and Belarus); (ii) with predominant I,J forms for **chemometrics**: Nordic countries (Sweden, Norway, and Denmark), Central Germanic countries (Germany, Austria, and The Netherlands), Romance countries (France, Italy, Spain, and Portugal), West Slavic countries (Poland, Czech Republic, and Slovakia), South Slavic countries (Slovenia, Croatia, Bosnia and Herzegovina, Serbia and Montenegro, Macedonia, and Bulgaria), selected Latin American countries (Brazil, Argentina, Mexico, and Chile), Finland, Greece, and Japan.

[a] Language codes are from Table III.

[b] Total number of scientific publications from the SCI, in thousands and rounded (the older search for Tot, Table II).

[c] The sum of relative frequencies contributing either to K (%K) or I,J (%IJ) end forms of the word **chemometrics** in national languages.

[d] Prediction of the end forms and their relative frequencies. Forms in '()' were predicted with lower or much lower frequencies than the dominant forms. The signs '+', '+–', and '–' mean good, partially good, and bad predictions, respectively.

[#] Estimated, from comparisons with the UK and Japan scientific productions.

3.4. Chemometrics in Europe: complete geographical pattern is visible

Europe is linguistically more homogeneous than Asia, with the Indo-European family predominant and other languages strongly affected by it (Finno-Ugric, Turkish, and Basque). The four Indo-European sub-families form six geographical units (Figure 1): Germanic, West Romance, West and East Slavic, South Slavic, East Romance (Romanian and Moldovan), and Baltic.

The last century was the time of great political changes, mostly located in Central and Eastern Europe. Countries and nations in these regions lived temporarily together in empires before the First World War (German, Russian, and Austro-Hungarian), and later in multinational states (Romania-Moldavia between the World Wars, USSR, Czechoslovakia, and Yugoslavia). All these facts as well as the new political and economic processes in Central and Eastern Europe and the increase of international collaborations (especially with English-speaking countries), have affected the science in Europe. Therefore, the forms for **chemometrics** (Table III) in some languages cannot be rationalized in terms of the considered science and extra-science determinants.

Data for **chemometrics** in Europe (Table III) account for 38 countries, 37 languages, 61 orthographic, and 76 pronunciation forms. Parts of the word **chemometrics** in national languages were analyzed separately, because of significant orthographic and pronunciation variations: **chem-** (Figure 5), **-mO-** (Figure 6), and **-TRIX** (Figure 7). Only the forms with significant frequency (over 10%, see Table II) were considered. There are general geographical patterns that

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Table V. Descriptors WWW and Pub and their geographic distribution[a,b,c,d]

Country (.xx)	Hits	Pub	Country (.xx)	Hits	Pub	Country (.xx)	Hits	Pub	Country (.xx)	Hits	Pub
USA (.edu, .gov, .us, .mil)	131.5	7.5	Romania (.ro)	0.2	0.2	American Samoa (.as)	0.03	0	Georgia (.ge)	<0.01	0
UK (.uk, .gb)	26.8	3.9	Tonga (.to)	0.2	0	Vietnam (.vn)	0.03	<0.1	Egypt (.eg)	<0.01	0.2
Germany (.de)	24.2	2.4	Western Samoa (.ws)	0.2	0	British Virgin Islands (.vg)	0.03	0	Guatemala (.gt)	<0.01	0
Netherlands (.nl)	18.6	1.2	South Africa (.za)	0.2	0.1	Lithuania (.lt)	0.02	<0.1	Lebanon (.lb)	<0.01	0
Denmark (.dk)	15.7	1.2	Slovenia (.si)	0.2	0.2	Moldavia (.md)	0.02	0	Morocco (.ma)	<0.01	<0.1
Sweden (.se)	13.5	1.7	Hong Kong (.hk)	0.2	0.3	Colombia (.co)	0.02	0	Sri Lanka (.lk)	<0.01	<0.1
France (.fr)	13.2	2.1	Turkey (.tr)	0.2	0.3	European Union (.eu)	0.02	–	Syria (.sy)	<0.01	0
China (.cn)	11.1	2.4	New Zealand (.nz)	0.2	0.1	Venezuela (.ve)	0.02	<0.1	Albania (.al)	<0.01	0
Italy (.it)	9.8	3.1	Israel (.il)	0.2	0.2	Saint Helena (.sh)	0.02	0	Bahrain (.bh)	<0.01	<0.1
Norway (.no)	9.7	1.0	Slovakia (.sk)	0.1	0.4	Peru (.pe)	0.02	<0.1	Bosnia and Herzegovina (.ba)	<0.01	<0.1
Spain (.es)	1.0	4.1	India (.in)	0.1	0.3	Ecuador (.ec)	0.02	<0.1	Christmas Islands (.cx)	<0.01	0
Japan (.jp)	1.0	1.2	Ireland (.ie)	0.1	0.4	Indonesia (.id)	0.02	<0.1	El Salvador (.sv)	<0.01	0
Australia (.au)	0.9	1.1	Croatia (.hr)	0.1	<0.1	Saint Tome and Principe (.st)	0.02	0	Macau (.mo)	<0.01	0
Brazil (.br)	0.9	1.4	Belize (.bz)	0.1	0	Turkey and Caois Islands (.tc)	0.02	0	Macedonia (.mk)	<0.01	<0.1
Belgium (.be)	0.9	1.2	Thailand (.th)	0.1	<0.1	Belarus (.by)	0.02	0	Palestine (.ps)	<0.01	<0.1
Poland (.pl)	0.9	0.9	Niue (.nu)	0.1	0	Cuba (.cu)	0.02	<0.1	Trinidad and Tobago (.tt)	<0.01	0
Tuvalu (.tv)	0.8	0	Chile (.cl)	0.1	<0.1	Seychelles (.sc)	0.02	0	Anguilla (.ai)	<0.01	0
Canada (.ca)	0.8	0.6	Mexico (.mx)	0.1	0.2	Ascension Islands (.ac)	0.01	0	Bangladesh (.bd)	<0.01	0
Russia (.ru)	0.7	0.3	Luxembourg (.lu)	0.1	0	United Arab Emirates (.ae)	0.01	<0.1	Brunei (.bn)	<0.01	0
Austria (.at)	0.6	0.5	Ukraine (.ua)	0.1	0.1	Kenya (.ke)	0.01	0	Djibouti (.dj)	<0.01	0
Czech Republic (.cz)	0.5	0.7	Singapore (.sg)	0.1	0.2	Uruguay (.uy)	0.01	<0.1	Lao (.la)	<0.01	0
South Korea (.kr)	0.5	0.2	Estonia (.ee)	0.1	0.1	Pakistan (.pk)	0.01	0	Liechteinstein (.ki)	<0.01	0
Taiwan (.tw)	0.5	0.1	Malaysia (.my)	0.1	0.1	Jordan (.jo)	0.01	0	Tanzania (.tz)	<0.01	0
Switzerland (.ch)	0.5	0.5	Latvia (.lv)	0.1	<0.1	Cyprus (.cy)	0.01	<0.1	Tunisia (.tn)	<0.01	
Finland (.fi)	0.4	0.4	Iceland (.is)	0.05	<0.1	Costa Rica (.cr)	0.01	<0.1	US Virgin Islands (.vi)	<0.01	0
Portugal (.pt)	0.4	0.5	Bulgaria (.bg)	0.05	0.3	Monserrat (.ms)	0.01	0	Zimbabve (.zw)	<0.01	0
Argentina (.ar)	0.3	0.4	Serbia and Montenegro (.yu, .cs)	0.05	0.1	Oman (.om)	0.01	<0.1	Faroe Islands (.fo)	0	<0.1
Hungary (.hu)	0.3	0.4	Saudi Arabia (.sa)	0.04	0.1	Mauritius (.mu)	0.01	0	Mali (.ml)	0	<0.1
Iran (.ir)	0.2	0.4	Cocos Islands (.cc)	0.04	0	Tokelau (.tk)	0.01	0	Rwanda (.rw)	0	<0.1
Greece (.gr)	0.2	0.4	Philippines (.ph)	0.03	0	Uzbekistan (.uz)	0.01	0	Sudan (.sd)	0	<0.1

[a] Country—countries and territories arranged in decreasing order of values of the descriptor WWW.
[b] (.xx)—WWW country domain.
[c] Hits—descriptor WWW (Table II) in thousands, rounded.
[d] Pub—descriptor Pub (Table II) in hundreds, rounded.

overcome linguistic and historical divisions as, for example, the (main) stress (') in **chemometrics**, which is located on the syllable **-me-** in **-'metrix** or at the end '**trix** in **-me'trix**. The distribution of the latter has a shape of the upside-down letter T (marked by solid boundary in Figure 7), with some minor fraction of **me'trix** in French and **'metrix** in Spanish-speaking areas. The **-mio-** forms in national languages exist only in West Romance but not in East Romance languages (due to long historical separation), and in Greek, whilst in all other language is **–mo–** with some pronunciation variations (Figure 6).

Swedish **kemometri** propagated into Denmark and Norway (**ke** → **kje**), and even to the Faroe Islands (high %IJ due to the collaborations with Danish Universities), but not into Icelandic. Iceland has its own word for chemistry (icl-1, without **chem-**), so an English orthographic form (**chemometrics**) was adopted. The Finnish equivalent **kemometri** + **-a** (the suffix for sciences in Finno-Ugric languages) was created and accepted in similar form in Estonia (which has 10% collaboration with Finland and 13% with Denmark, Norway, and Sweden). Dutch has also its own word for chemistry (see dut-1, Table III), and this form **Chemometrie** was created also in the German-speaking area. Afrikaans, originated from 18th century Dutch dialects, is today linguistically distinct and geographically far from Dutch. Two forms for **chemometrics** (afr-1 and afr-2) were created in Afrikaans. This is a unique example where a language takes distinct K forms that differ only in the final

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

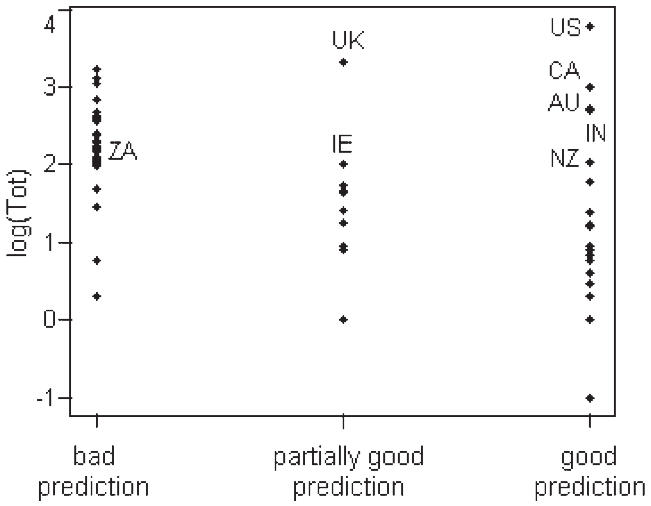


Figure 4. Relationship between K/I,J end form prediction for chemometrics (Table IV) and the total number of scientific publications Tot.

letter and sound (**-k** and **-ka**). The French dominant forms **chimiométrie** (fre-1) and **chimiometrie** (fre-2), coming from the word for chemistry (**chimie**), have their analogs on the Iberian Peninsula. The dominant Italian and Romanian forms do not follow their words for chemistry, but are more similar to the international form **chemometrics** (**chem-** instead of **chim-**, Figure 5). The Moldavian form (mld-1: **chemometrica**) was created not as Romanian forms (as one would expect), but like the dominant form **khemometrika** in the East Slavic languages, probably due to intensive collaboration with Ukraine (7%, Table IV). The Albanian Tosk form (alt-1) was correctly predicted (Table IV) as I,J type. It was not created from **chemistry** (**kimia**), but probably from the Italian form, because Italy is the major traditional partner of Albania in scientific collaborations (25%). However, **chemistry** in Albanian Gheg (the official language in the major part of the Kosovo region in Serbia) is different (**kimi**), and hence, a new form for **chemometrics** could be expected in this language (**kemometri**).

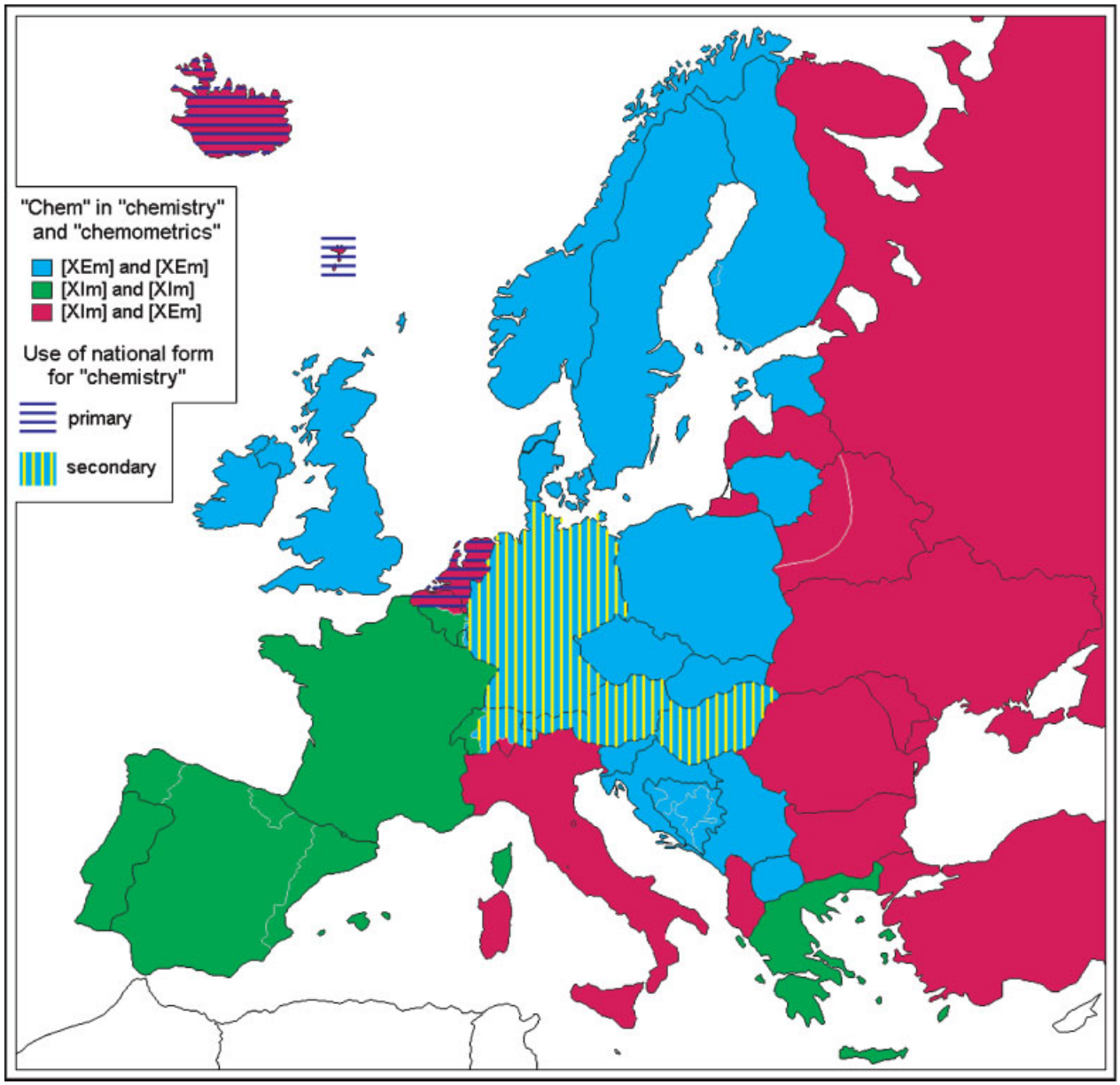


Figure 5. Comparison of the combining forms **chem-** in **chemometrics** and **chemistry** in official languages in Europe.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

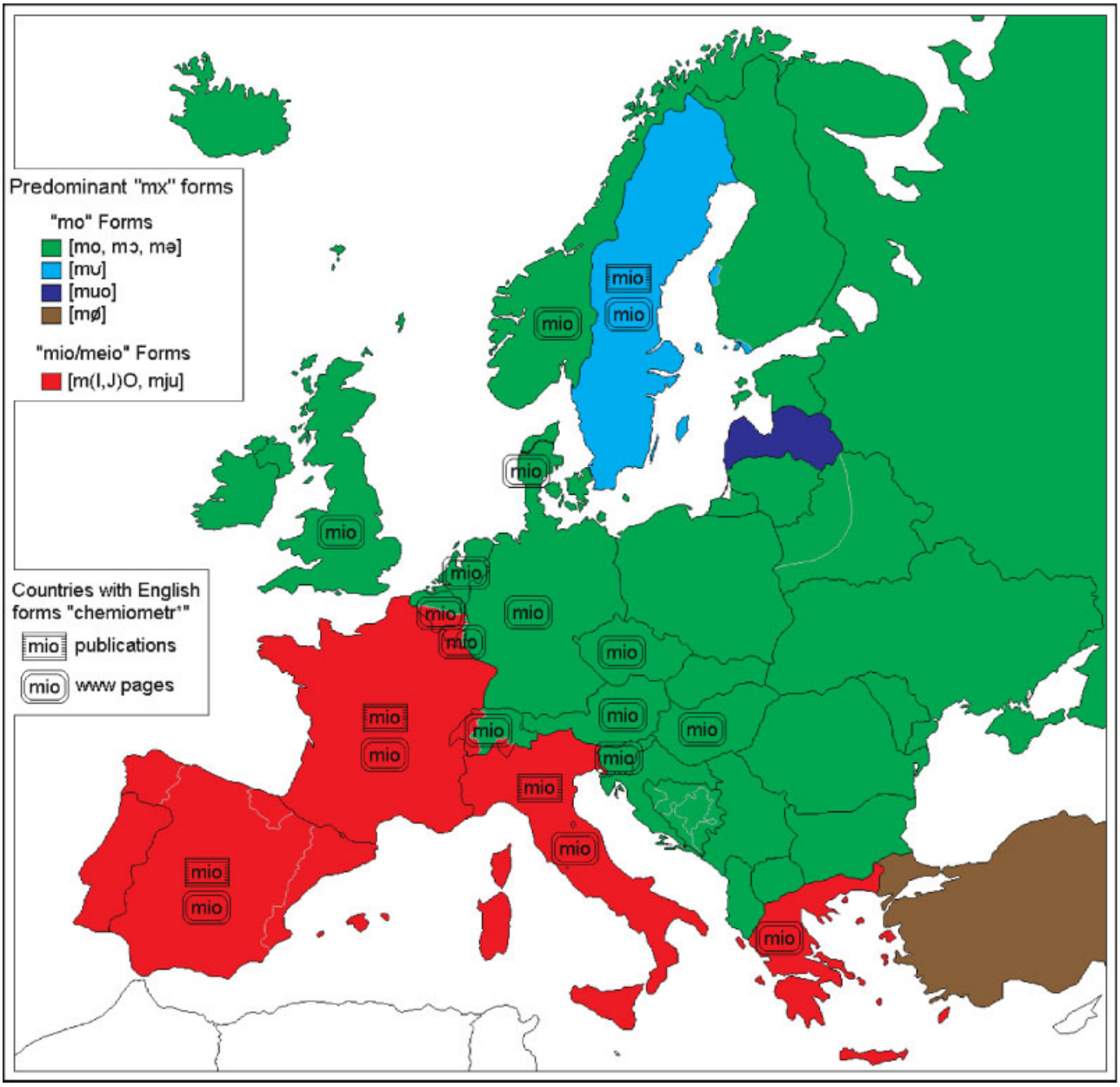


Figure 6. Syllable **-mO-** in **chemometrics** in official languages in Europe. The presence of **-miO-** forms on the internet and in SCI publications is also marked.

Most East European countries did not coin **chemometrics** from their **chemistry** (Figure 5). The dominant Russian **khemometrika** (K type, rus-1), with **khem-** and not **khim-**, agrees with the international (English, K type) form. This form is also used in other former USSR republics with significant number of native Russian speakers (Ukraine, Belarus, Latvia). Other Slavic countries, partially Belarus and Ukraine, and Lithuania adopted I,J types for **chemometrics** as did their western neighbors. Some Slavic countries remained divided between the K (similar to the international standard) and I,J forms (Czech Republic, other West Balkans, Figure 7). The geographic pattern of all orthographic and pronunciation end types **TRIX** together with their frequencies are shown in Figure 7. Three Latvian forms, among which two start with **kem-** (lav-2 and lav-3) and one with **hem-** (lav-1) and not with **şim-** as in **şimija** (**chemistry**), were positively predicted (Table IV) due to significant international collaboration of Latvia with diverse country groups: Central Germanic, English-speaking countries, Nordic, East Slavic, and West Romance countries.

3.5. Pluralism in creation of forms of chemometrics in national languages

Five mechanisms are responsible for the pluralism in orthography and pronunciation of **chemometrics**. The first is of etymological nature: to create a K or I,J form. This K/I,J distinction comes from Classical Greek/Latin [8,74,75] words: substantives **metria** (feminine: measurement, measuring), and adjectives **metrikos/metricus**, or **metrike/metrica** or even **metrikon/metricum** (masculine,

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

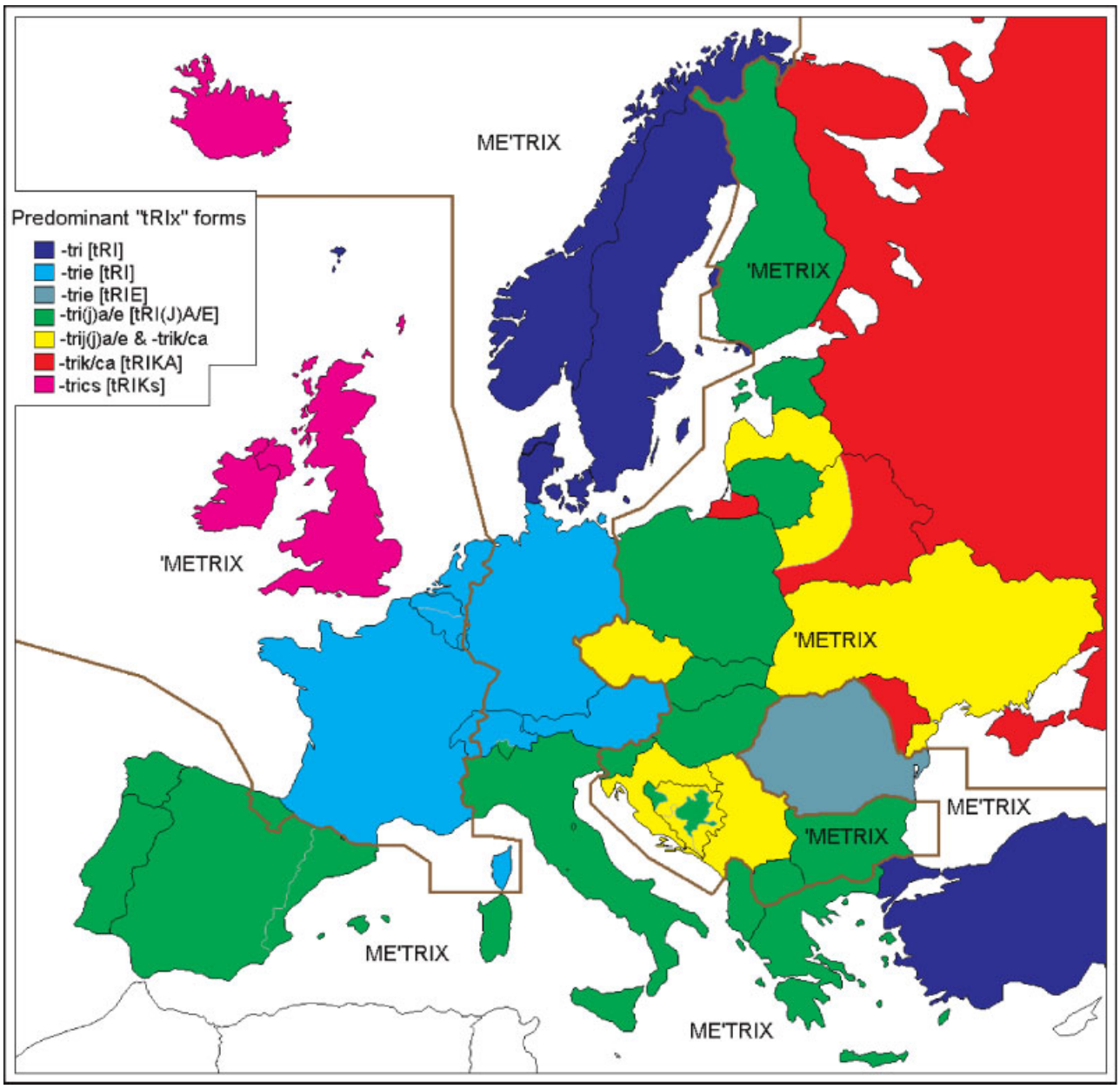


Figure 7. Orthographic **TRIX** and pronunciation [TRIX] end forms and the stress position in **chemometrics** in official languages in Europe. Putative solid boundaries delimit the two stresses.

feminine or neuter, respectively: in meter, by meter, metric). All these words have their origin in **metron**/**metrum** (substantive: measure, tool for measuring) or **metrein**/**metiri** (transitive verb: to measure). The Greek forms and Latin **metiri** came from the proto-Indo-European language (PIE), the ancestral language of all Indo-European languages, talked about 3000–4000 years BC, right before the Indo-European sub-families branched (Scheme 1) [8,74–76]. Reconstructed PIE roots for the transitive verb *to measure* (**med-**, **met-**, **mer-**, **meəl-**, etc.) can be recognized in many extinct and living Indo-European languages. The existence of similar roots and words has aided the spread of scientific names ending with **-metry** or **-metrics** in these languages.

Modern English spoken by native speakers distinguishes well the meanings of the feminine names of '*-y*' and '*-ics*' sciences. The suffix composite forms **-metry** and **-metrics** denote 'the process or science of measuring' and 'the application of statistics and mathematical analysis to a specified field of study', respectively [74]. As S. Deming said in an interview [14], **chemometrics** is an appropriate name given to *collected* mathematical and statistic methods in chemistry. The first '*-y*' sciences and, in particular, '*-metry*' sciences, date from Classical Greek. For example, **geometry** had the following linguistic history as a feminine substantive in singular [8,9,74,77,78]: geome*tria* (Classical and Late Greek) → geome*tria* (Classical, Late, and Medieval Latin) → géomé*trie* (Old and Middle French) → geome*trie* (Middle English) → geome*try* (Modern English). This example explains the **TRIX** in **chemometrics** as found in all West Romance and most Germanic languages (the I,J

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

end forms, see Table III and Figure 3). The word **metrics** and the first *-metrics* sciences are not more than 100 years old, although other '*-ics*' sciences date from the Renaissance. There is obvious similarity between these '*-metrics*' feminine substantives and the corresponding adjectives in Indo-European languages: English chemometr*ics*–chemometr*ic*, German: Chemometr*ik*–chemometr*isch*, Croatian kemometr*ika*–kemometr*ička*, Czech chemometr*ika*–chemometr*icka*, Danish kemometr*ik*–kemometr*isk*, etc. This is because **metrics** and older '*-ics*' scientific names were formed from their adjectives, and newer names adopted the same '*-ics*' shape. These transformations were introduced in English 'to mark the *complex nature* of such sciences and it may have been an imitation of the use of the Greek plurals' [77]. The most probable evolution of English **metrics** [8,9,74,77,78] started with Classical Greek adjectives, feminine singular *metrike* or neuter plural *metrikea*, continued with their Classical-Late-Medieval Latin equivalents of the same form *metrica*, and through Old or Middle French *métrique* ended in Middle English *metric*, from which modern English created the adjectives *metric* and *metrical* and the substantive *metrics*.

The second mechanism of the **chemometrics** pluralism in countries with relatively modest scientific production (Table IV) may lie in foreign influences, as has been shown in the previous section. Figure 8 illustrates how the number of the orthographic forms for **chemometrics,** as well as the frequency of the most dominant forms, are distributed in Europe. The places where different languages or even language groups meet are well visible in Latvia, at the bordering areas of the French-speaking territory, and the cosmopolitan city of London.

The third mechanism of the **chemometrics** pluralism is related to countries with relatively high scientific production,

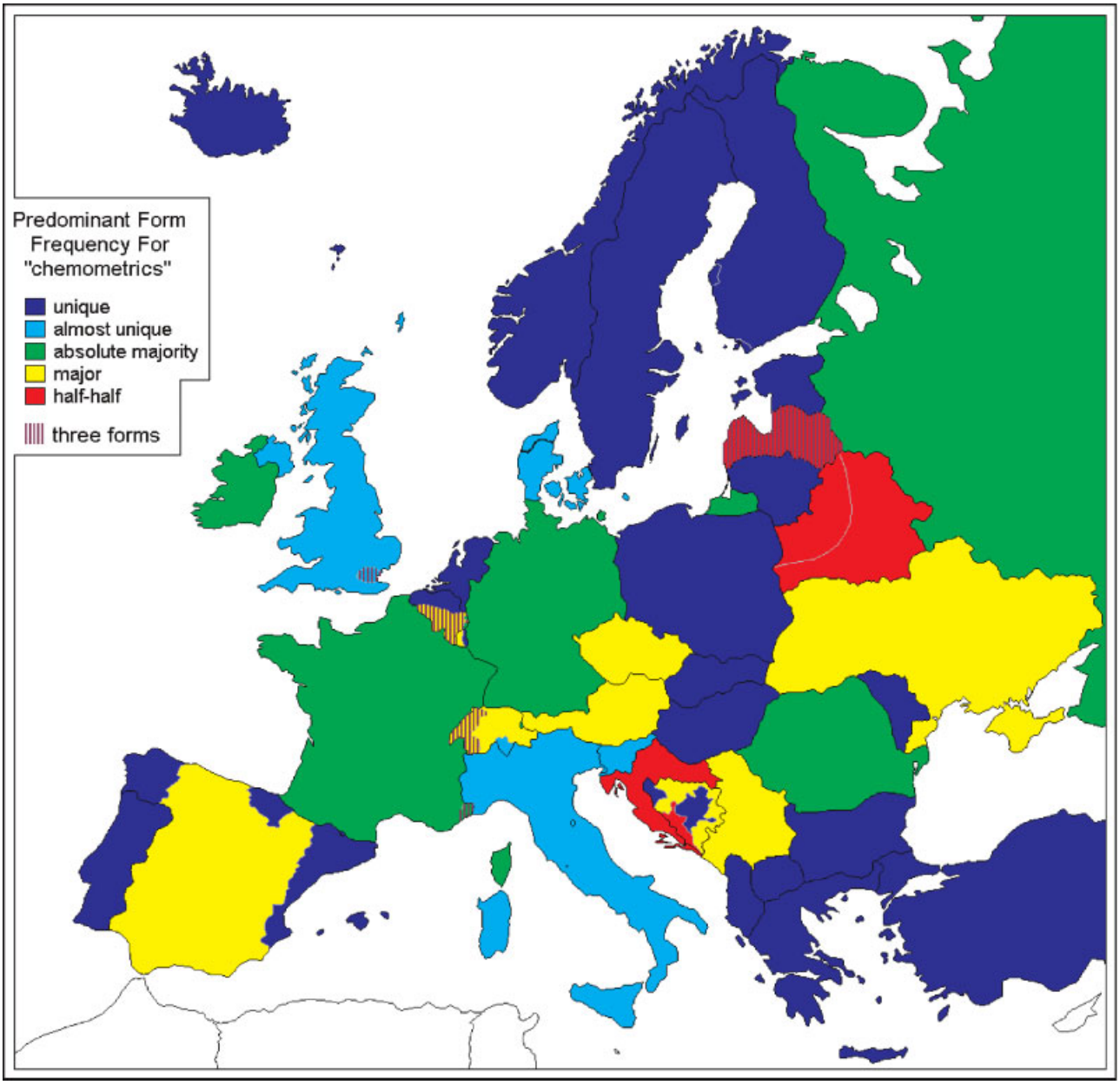


Figure 8. Frequency of the predominant form(s) and the number of existing forms of **chemometrics** in official languages in Europe.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

large territories and significant populations. Such countries usually have more than one language standard and consist of regions with different linguistic preferences for certain terms. For example, China and Japan have four written forms for **chemometrics** (Table III).

The fourth mechanism of the **chemometrics** pluralism includes large language areas that extend over two or more countries or even continents and show much linguistic diversity. China uses forms for **chemometrics** written with Traditional, Modern, or mixed Chinese characters (cmn-1 to cmn-3), but Taiwan prefers not to use the mixed combinations (cmn-3). In the German-speaking area, the minor form **Chemometrik** is more frequent in Austria and Switzerland than in Germany. Russian **khemometriya** (rus-2) is more frequent in Ukraine and Belarus than in Russia. French also shows differences between the European and Québec variants, from which the latter has only two forms (fre-1 and fre-2). Figure 8 shows **chemometrics** pluralism in Europe, with countries forming groups with the same pluralistic characteristics, and also at the meeting points of these groups.

The fifth mechanism of the **chemometrics** pluralism is related to the use of English. The English language represents a special case because it is the most frequently and universally used language in science, and includes many non-native English speakers. These speakers also contribute to the development of scientific English. Six orthographic forms for **chemometrics** (eng-1 to eng-6, Table III) were found in internet documents and four of them (eng-1 to eng-4) in the WOS. Adjectives **chemometric/chemometrical**, **chemiometric/chemiometrical**, and **chemimetric** were also found online and in the WOS. **Chemometry** is a singular noun for which two inflected forms were detected: plural **chemometries** (online and in the WOS) and Saxon Genitive **chemometry's** (online). Rare forms **chemimetrics/chemimetry** were mostly found at UK and USA sites. With respect to the other four forms, English-speaking countries strongly prefer **chemometrics** and the adjective **chemometric**, especially USA and UK. It is clear that the forms **chemometry**, **chemiometry**, and **chemimetrics** and their adjectives are preferred by non-native English speakers from countries where the most dominant forms are NOT of the K type, *that is*, do not end with *-ics*, *-ik*, *-ika*, etc. The main users of these I,J forms are Western Romance countries and Brazil, followed by Central Europe and Scandinavia (Figure 6), China, and Japan. For the East Slavic countries, such forms have not been found, which is quite surprising. It seems that the tendency for their use has its roots in the first languages of the speakers: **-mio-** syllables in West Romance languages and Greek, **-metri/ie/ia/ija** ends in most West and Central Europe and Scandinavia. **Chemometry** and **chemometrical** are more frequent in Western and Central Europe than in Eastern Europe (Figure 9).

3.6. Chemometrics all around the world

Table V lists the countries and territories for which non-zero Google hits for English **chemometrics** (eng-1) were found. A considerable number of islands and small countries have domains with advertising material, copies of pages from other sites, and lists of journals, where **chemometrics** appears. In total, 115 from 250 (46%) countries and territories are in some way related to chemometrics, which accounts for about 65% of the world land area and 86% of its population.

3.7. Chemometric activity may be related to the usage of the word chemometrics in publications and internet documents

The 14 bibliometric and webometric (internet-based) descriptors as well as corresponding analyses should be understood in a qualitative and general sense, with no intention for country-to-country or language-to-language comparisons of a competitive nature. Generation of more exact chemometric activity descriptors is an impossible task because it should take into account all kinds of activities that are recorded in national languages published by all electronic and classical means.

WWW, the Google hits for **chemometrics** in national languages (Table II), varied from 1 to more than 100 000 (Table V). The situation in Europe (Figure 10) is geographically and historically determined. Most hits come from Germanic and West Romance languages (over 1000 hits per country), with some exceptions (Switzerland, Austria, Belgium, Iceland, Faroe Islands, Luxembourg). In the world, only USA, China, and Brazil have more than 10 000 hits, and Japan 1000–10 000. Moderate numbers of hits, *that is*, 100–1000, was recorded for other English-speaking countries, Argentina, and some Asian countries (The Philippines, Taiwan, India, and Singapore). The countries with 10–100 Google hits are concentrated in Asia (South Korea, Thailand, Turkey, Malaysia, and Indonesia), Latin America (Mexico, Cuba, Colombia, Chile, and Peru), and Kenya. Less than 10 hits were found in other Asian countries (see Figure 2), other Latin American countries, and a few African countries (Table V). One should bear in mind that these numbers represent complex phenomena as in the previous section. Besides chemometric activity, population size, commercial advertisements, bibliographic records, and copies of texts from other sites determine the number of internet documents containing **chemometrics**.

The numbers of SCI publications with **chemometr*** in the address and in topics (Pub descriptor, Figure 11 left) show an exponential-like increase with time, meaning that chemometric groups and others using chemometric methods are multiplying and becoming more productive. However, the total number of considered publications (Pub = 3858 for the world) makes up some 20% of all chemometric publications in the WOS (about 15 000 were reported in 2004 [79]). The same percentage is obtained from the number of Iranian publications (about 200, from which many are not reported in the SCI [80]) compared to the number in Table V (40). The descriptor Pub can be represented as a distribution function, a function of time and the number of participating countries. The number of countries depends on the class to which they belong (Figure 11). Class means the range of the number of publications expressed in log form. The distribution function for the world (Figure 11 middle) exhibits a tendency to become Gaussian in the near future (two

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

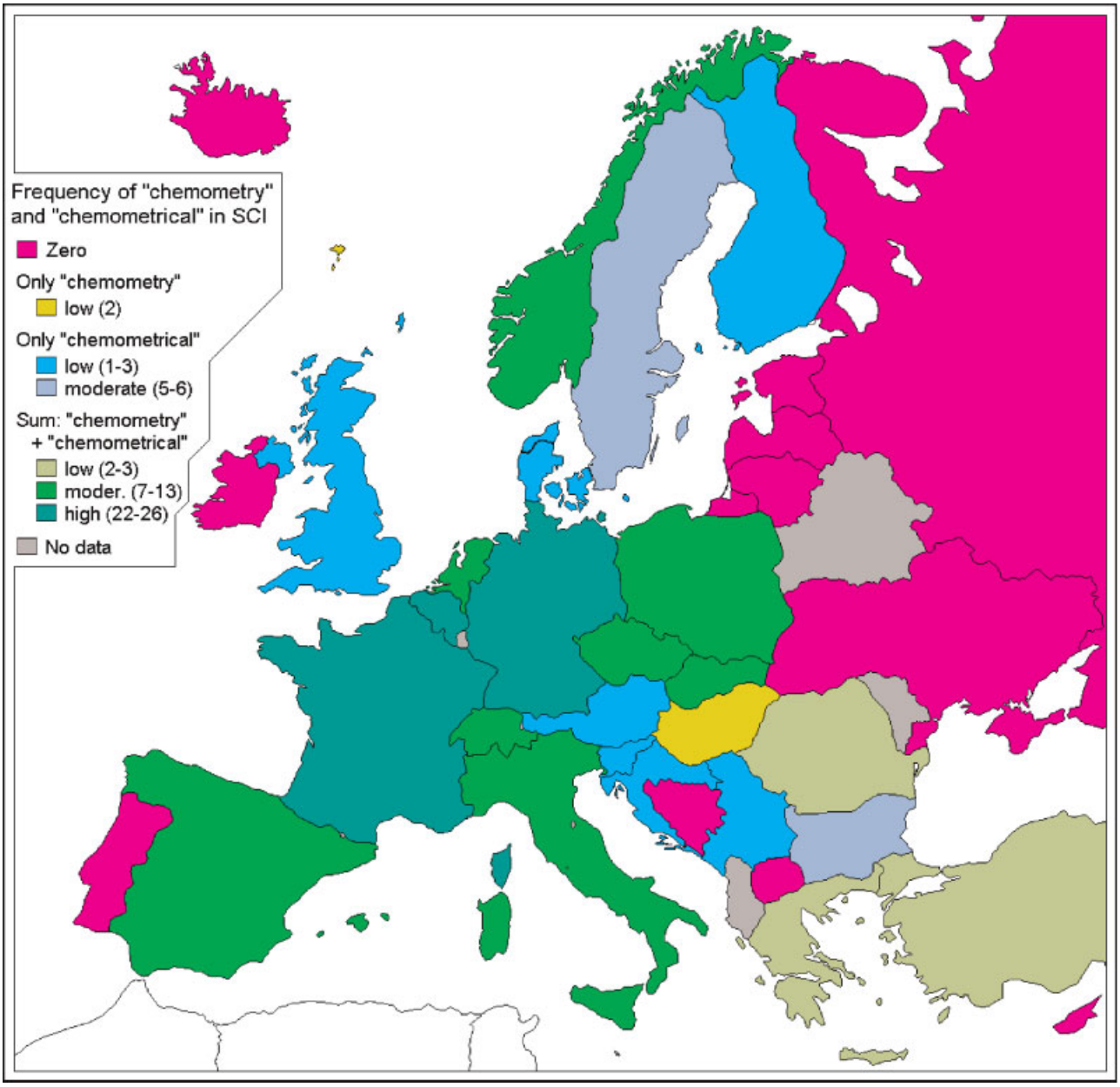


Figure 9. Frequency of the English forms **chemometry** and **chemometrical** in SCI publications in Europe. The number of corresponding publications is cited in the brackets in legends.

decades from now). To achieve a normal distribution for all existing countries and territories, we should wait for the 100th birthday of the name **chemometrics**. Other continents (results not shown, except for Europe in Figure 11 right) show similar, although less pronounced trends. These bibliometric tendencies reflect the differences between developed countries and countries in development, as also noticed by Ojasoo [66]. Europe is very peculiar, because it represents a high concentration of countries chemometrically very active, and development differences are rather moderate. The formation of a Gaussian-like distribution curve is well visible (Figure 11 right). A hypothetical situation (Europe with USSR, Czechoslovakia, and Yugoslavia in 2005, dashed curve), when compared to the reality, shows that the Gaussian curve formation is slowed down due to political and socio-economic changes in Central and Eastern Europe [81] that affected the science in these countries.

3.8. Chemometrics-development relationships

Besides Pub and WWW, other bibliometric and webometric descriptors (Table II) can be directly or indirectly related to chemometric activities of 36 countries in Europe and 76 countries worldwide. The correlation coefficients between these descriptors range from 0.56 to 0.96 for the Europe data set and from 0.42 to 0.99 for the world data set. Some

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

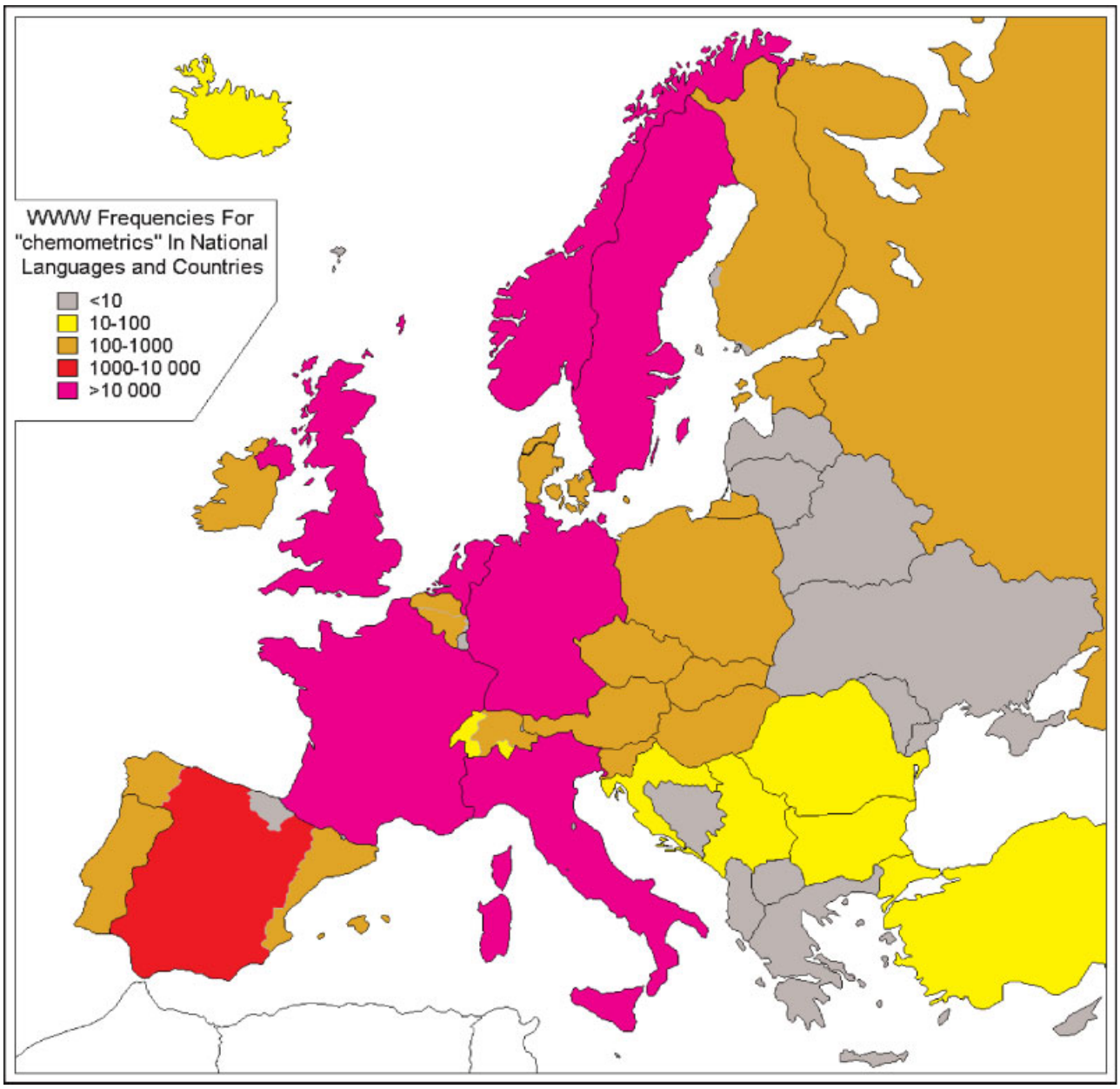


Figure 10. Number of Google hits for **chemometrics** in official languages in Europe.

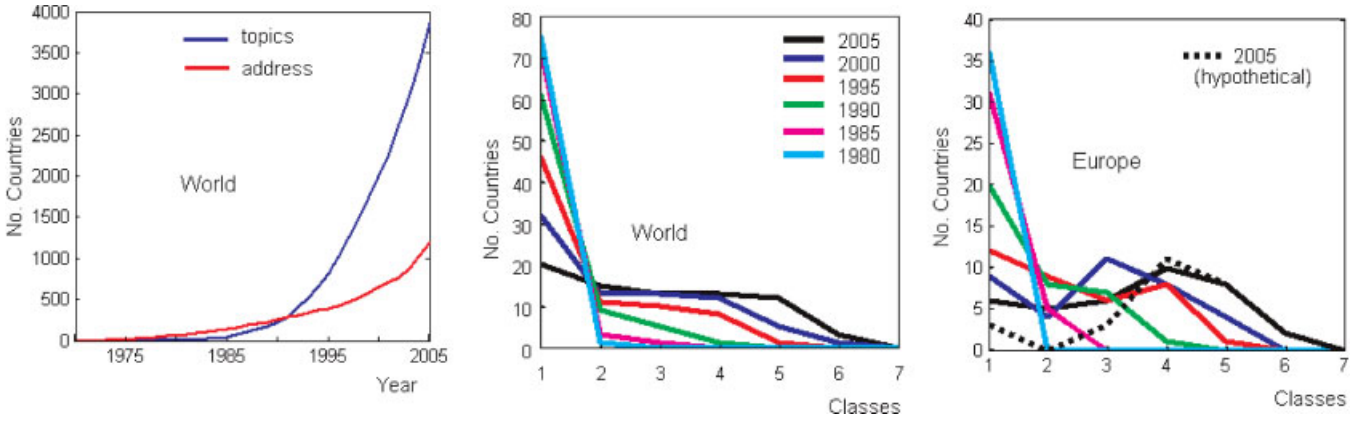


Figure 11. Left: Number of the SCI publications with **chemometr*** in topics (descriptor Pub) and addresses represented as a function of time. Middle and right: Distribution function of the descriptor Pub. The number of countries is a function of time (different colors) and class. Class is defined as the range of the number of publications in 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). The curves for European countries, except for the dashed curve, include countries that became independent after the disruption of the USSR, Czechoslovakia, and Yugoslavia.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

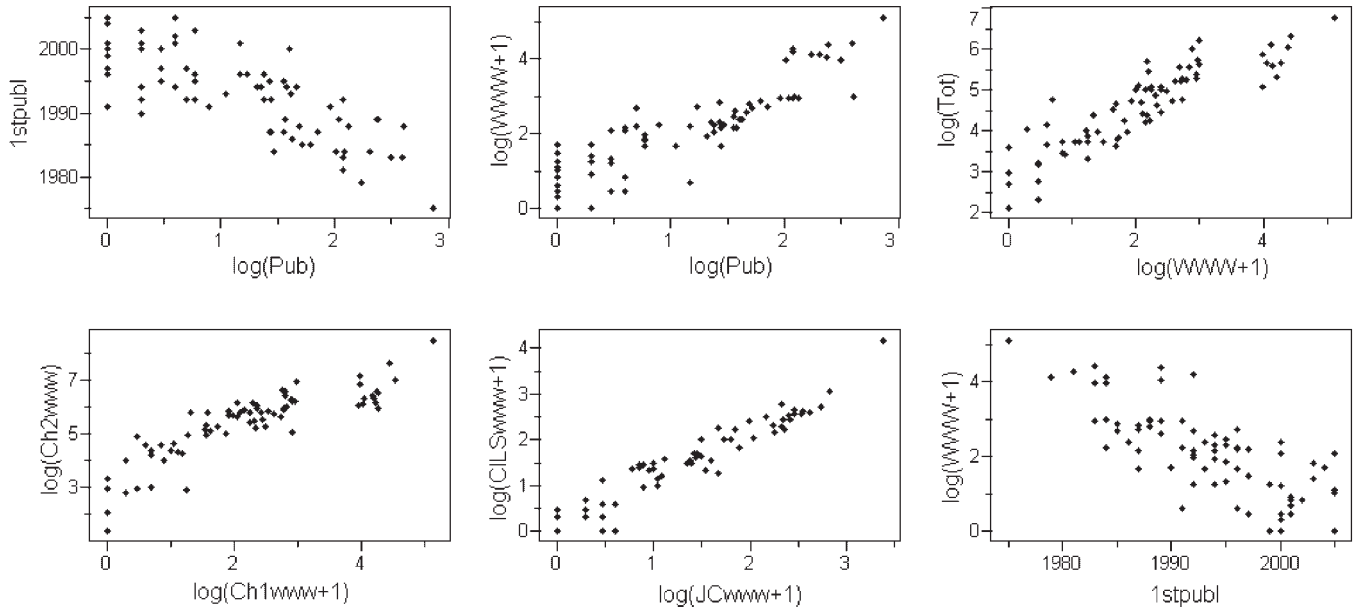


Figure 12. Scatter plots showing representative correlations between chemometric activity descriptors for the world data set.

representative correlograms (Figure 12, the world data set) show that more intense chemometric activity is related to longer chemometric tradition (1stpubl), SCI publishing (Pub, Tot), and web publishing (WWW and other descriptors). The following direct relationships can be observed: between bibliometric and webometric descriptors, between chemistry and chemometrics (Ch2www and Ch1www), and between chemometrics and scientific production (WWW, Tot). Google hits for the two chemometric journals (JCwww and CILSwwww) are mutually well correlated, and besides chemometric research may include literature citations and library documents containing the journals' names.

Correlation coefficients between the development indices range similarly for the Europe (0.54–0.96) and world data sets (0.48–0.98). These indices correlate weakly to moderately with chemometric activity descriptors, with correlation coefficients being 0.4–0.67 and 0.15–0.66 for the Europe and world data sets, respectively. The representative correlograms in Figure 13 show rather obvious relationships between some development indices (policy, diffusion, access, and DAI) and scientific (Tot), chemical (Ch2www), and chemometric (Chempubs) publishing. In spite of noticeable dispersion, the general tendency can be observed: higher index values are related to higher chemometric

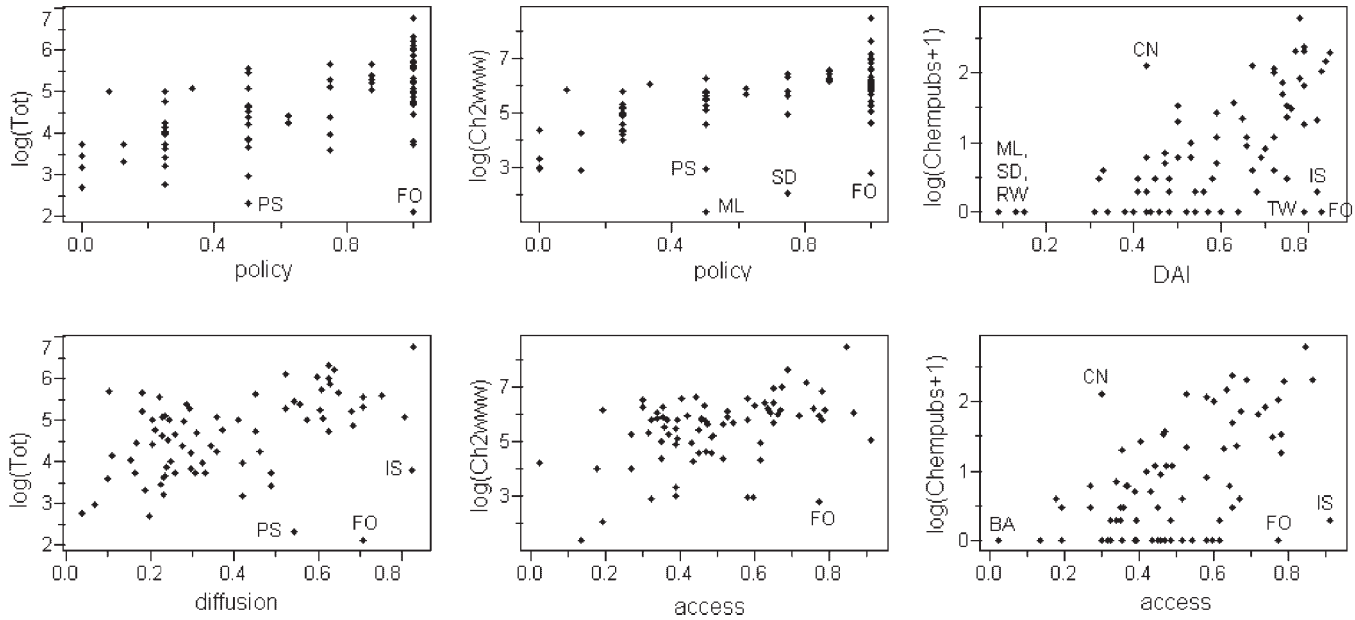


Figure 13. Scatter plots showing representative correlations between chemometric activity descriptors and development indices for the world data set. Countries which deviate from the main trends are marked with domain names as in Table V.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

activity. This fact may be related with earlier observations [66–68] that national scientific efforts are well correlated with scientific production.

PCA and HCA with complete linkage method were performed for both Europe and world data sets due to correlations between the 22 descriptors. The Europe data set is well characterized by two principal components when leave-one-out crossvalidation is applied (PC1: 65%, PC2: 20%, and PC3: 3% variance, with root mean square error of cross validation RMSECV 2.95, 2.05, and 1.94, respectively). The biplot in Figure 14 was constructed as described by Geladi *et al*. [82] using the autoscaled data matrix **X** (meancentered and scalled to unit variance). **X** was decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}^{1/2}\,(\mathbf{S}^{1/2}\mathbf{V})^{\mathrm{T}} + \mathbf{E}$ where **U** and **V** are the orthonormal scores and loadings matrices, respectively, and **S** and **E** are the singular values and residuals matrices, respectively. Figure 14 shows noticeable clustering and internal cluster structures that can be noticed also in the HCA (not shown) and can be related to the geographic, historical, and linguistic patterns already discussed in previous sections. More intense chemometric activity is related mainly to more negative PC1, where development indices (**16**–**22**)

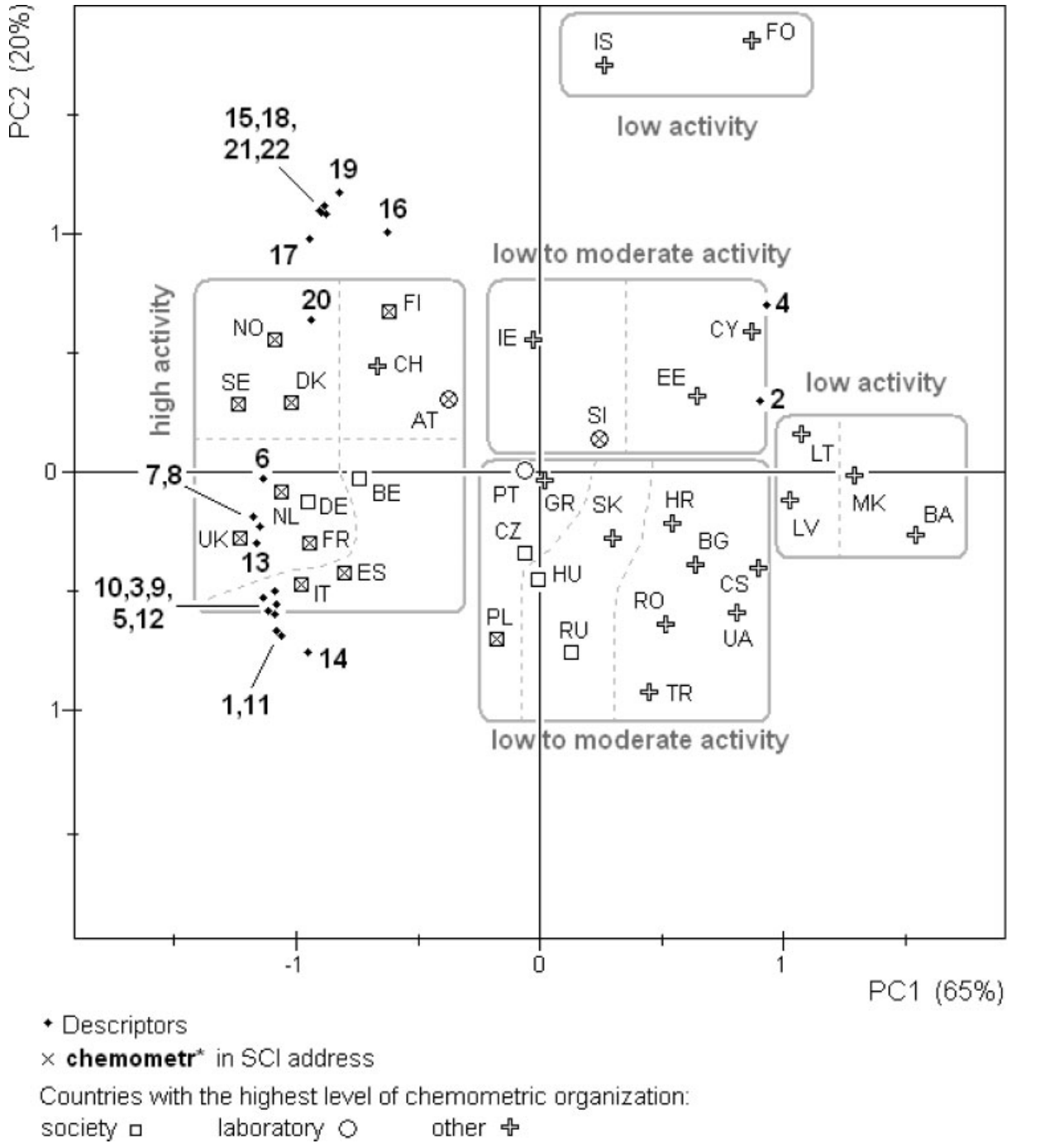


Figure 14. PCA biplot for the Europe data set (marked with country domains as in Table V). The highest organization levels of chemometrics and the presence of SCI publications with **chemometr*** in address are marked by different symbols. HCA clusters and sub-clusters are drawn in gray tones. The descriptors are marked with numbers as in Table II.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

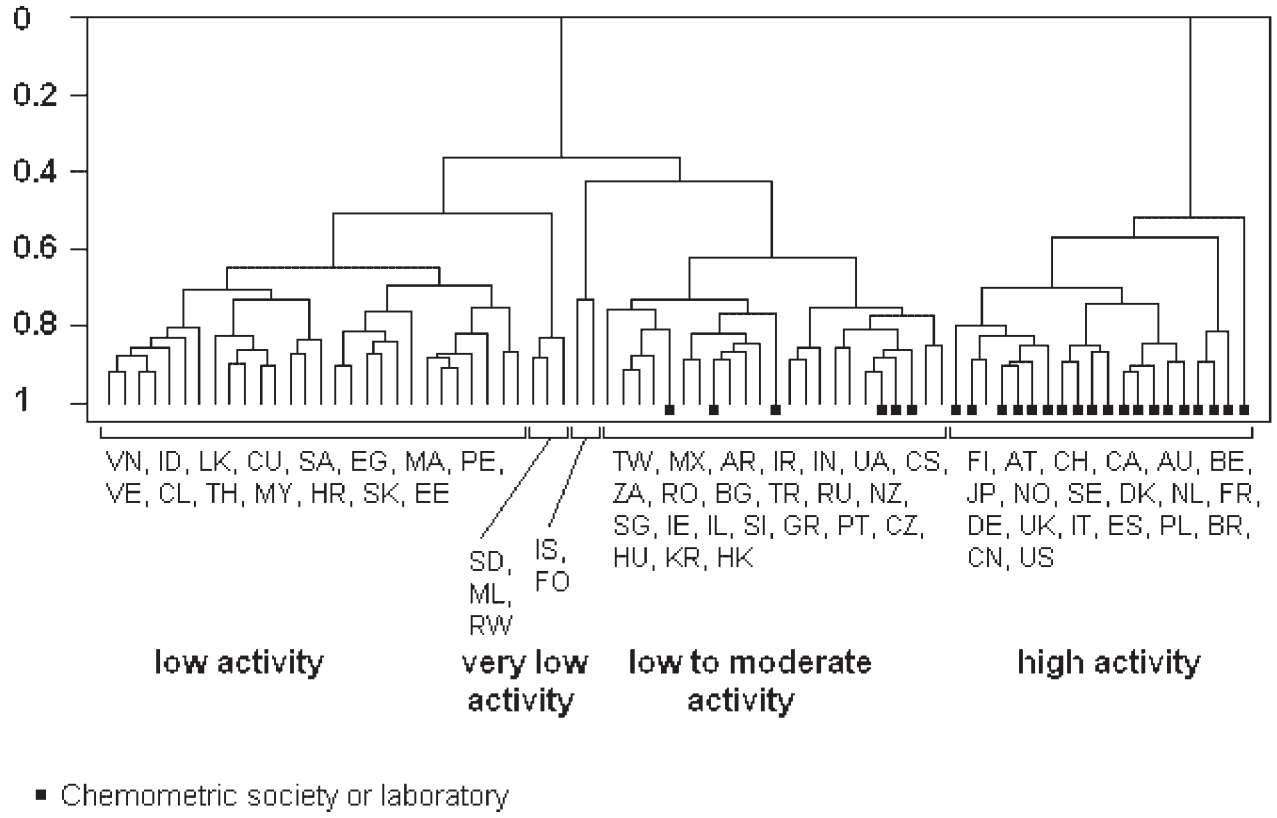


Figure 15. HCA dendogram for the world data set, with clusters of different chemometric activity levels. Countries are marked by their domains as in Table V. The existence of chemometric society or laboratory is marked by solid squares.

and descriptors related to numbers of publications (**1**, **3**, **5**–**14**) have more weight and descriptors related to the date of the first publishing (**2**, **4**) have low weight, *that is*, they indicate longer publishing tradition. PC2 shows mild discrimination with respect to the development indices and publishing descriptors. Countries with chemometric societies and laboratories are mainly concentrated in the high activity cluster and partially in the two neighboring clusters. The world data set is a qualitative extension of the Europe data set when PCA is performed, and being described by two PCs (PC1: 68%, PC2: 14%, and PC3: 4% variance, with RMSECV 2.73, 2.10, and 1.92, respectively). The corresponding HCA dendogram (Figure 15) exhibits interesting cluster structure with respect to the level of chemometric activity. The countries with chemometric societies and laboratories and **chemometr*** in addresses of SCI publications are concentrated in the high activity cluster and partially in the moderate activity cluster. These countries are generally more developed and make great efforts to invest in science and technology. Based on these exploratory analyses, it is possible to conclude that the descriptors based on the word **chemometrics** as measures of chemometric activity showed rather clear connection between chemometrics and scientific and technological development.

4. CONCLUSIONS

The subject of the present work was the word **chemometrics** in English and national languages from etymological, linguistic, and bibliometric points of view. Internet and literature searches resulted in 82 orthographic and 127 pronunciation forms of **chemometrics** in 48 languages belonging to 10 language families and language isolates, distributed over all continents.

The orthographic end forms of the '*-y*' and '*-ics*' types as well as other fragments of **chemometrics** show interesting distribution patterns in Europe and Asia, significantly determined by international scientific collaboration, the country's development and other extra-science factors such as geography, history, and language. Complex classification of these end forms into at least three groups (I, J, and K) rather than in only two (I and K) is also demonstrated.

Correlation, Principal Component, and Hierarchical Cluster Analyses of the Europe and world data sets described by 22 bibliometric, webometric, and country development indices show that intensive chemometric activities are generally related to considerable scientific production, chemometric tradition, and satisfactory scientific and technological development. These indices may indicate the existence of chemometric laboratories and societies in most countries under investigation. Europe and North America are still the primary chemometric regions. Other English-speaking countries, China, and Japan are also strong chemometric regions. New chemometric regions are arising in Latin America and throughout Asia. Africa, with the exception of South Africa, is still far from notable chemometric activity. It is expected that more homogeneous and normally distributed geography of chemometrics in the world would be achieved in a rather long time (70 years from now).

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem

Acknowledgements

The authors acknowledge The State of São Paulo Funding Agency (FAPESP) for financial support and Dr Carol H. Collins for English revision.

REFERENCES

1. Partington JR. *A Short History of Chemistry (2nd edn)*. MacMillan: London, 1951; 20–25.
2. Salzberg HW. *From Caveman to Chemist: Circumstances and Achievements*. American Chemical Society: Washington, DC, 1991; 33–34.
3. Hudson J. *The History of Chemistry*. McMillan: London, 1992; 22.
4. Mierzecki R. *The Historical Development of Chemical Concepts*. Kluwer: Dordrecht, The Netherlands, 1991; 6–7.
5. Moore FJ. *A History of Chemistry (3rd edn)*. McGraw-Hill: New York, NY, 1939; 2–3.
6. Multhauf RP. *The Origin of Chemistry*. Franklin Watts: New York, NY, 1966; 102–105.
7. Maar JH. *Pequena História da Química*. Papa-Livro: Florianópolis, SC, Brazil, 1999; 25–26.
8. Online Etymological Dictionary. http://www.etymonline.com/index.php [24 May 2006].
9. Merriam-Webster Online Dictionary. http://www.m-w.com [24 May 2006].
10. Chemometr-ics/y. A message posted by Svante Wold at ICS-L@listserv.umd.edu on February 24, 2006.
11. Forty aspects of Umeå University. http://www.umu.se/infoenheten/jubileum05/images/pdf/jubskrift_eng.pdf [18 February 2006].
12. Kowalski BR, Brown S, Vandeginste B. Editorial. *J. Chemometr.* 1987; **1**: 1–2.
13. Announcements: Chemometrics 15 year, a Birthday Celebration Symposium Västerbacken Kursgård, Holmsund (Umeå), Sveden, 7–11 June 1989; *J. Chemometr.* **3**: *453.*
14. Esbensen K, Geladi P. The start, early, history of chemometrics: selected interviews. Part 2. *J. Chemometr.* 1990; **4**: 389–412.
15. Wold S. Spline-funktioner-ett nytt verktyg i data-analysen. *Kemisk Tidskrift* 1972; **3**: 34–37.
16. Wold S. Nonsymmetry of reaction parameter (ρ) and substitution parameter (σ) in Hammet equation and similar extrathermodynamic relationships. *Acta Chem. Scand.* 1973; **27**: 3602–3604.
17. ISI Web of Knowledge database of the Thompson Corporation. http://portal.isiknowledge.com/portal.cgi [6 June 2006].
18. Wold S. Chemometric: what do we mean with it, and what do we want from it? *Chemometr. Intell. Lab. Syst.* 1995; **30**: 109–115.
19. Kowalski BR. Chemometrics: views and propositions. *J. Chem. Inf. Comput. Sci.* 1975; **15**: 201–203.
20. Kowalski BR (ed). *Chemometrics: Theory and Applications*, ACS Symp. Ser 52. American Chemical Society: Washington, DC, USA, 1977.
21. Chemometrics in Analytical Chemistry: proceedings of a conference held in Petten, The Netherlands, September 15–17, 1982. *Anal. Chim. Acta* 1983; **150**.
22. Jacobsen T, Volden R, Engan S, Aubert O. Chemometric study of some beer flavor components. *J. Inst. Brewing* 1979; **85**: 265–270.
23. Kowalski BR. Chemometrics. *Anal. Chem.* 1980; **52**: 11R–122R.
24. Vandeginste B, Kowalski BR. Training chemometricians at the University of Washington and the University of Njimegen. *Abstr. Pap. Amer. Chem. Soc.* 1981; **182**: 12-CHED.
25. Buyukkoca E, Yamada I. A new approach to regression analysis in chemometrics. *Abstr. Pap. Amer. Chem. Soc.* 1984; **188**: 57-ENV.
26. Jinno K, Ohta H, Hirata Y, Sasaki S, Abe H. Microcolumn liquid-chromatography combined with computer-assisted retention prediction system for polycyclic aromatic hydrocarbons in extract from diesel particulate matter. *Anal. Lett. A* 1984; **17**: 905–913.
27. Bek WK. Colloquium Chemiometricum Mediterraneum, an international forum for chemometrics, Barcelona, Spain, 9–11 November 1987. *Chemom. Intell. Lab. Syst.* 1988; **5**: 98.
28. Tomàs X, Casassos E. Editorial. *J. Chemometr.* 1988; **3**: 101.
29. Esbensen KH. First Scandinavian Symposium on Chemometrics (SSC1), Lappeenranta, Finland. 6–8 October 1988. *Chemom. Intell. Lab. Syst.* 1989; **6**: 2–4.
30. Geladi P, Esbensen K. The start, early, history of chemometrics: selected interviews. Part 1. *J. Chemometr.* 1990; **4**: 331–354.
31. Kiralj R. Chemometrics in different languages, Laboratório de Quimiometria Teórica e Aplicada, UNICAMP – Campinas SP. http://pcserver.iqm.unicamp.br/~rudolf/chemometrics.html [4 September 2002].
32. Faber K. Chemometry Consultancy: Chemometrics. http://www.chemometry.com/Index/Chemometrics.html [30 March 2006].
33. Chemometry/-ics. A message posted by Klaas Faber at ICS-L@listserv.umd.edu on February 24 2006.
34. Basic tasks and goals of the Russian Chemometrics Society (in Russian), at the page of the Russian Chemometric Society. http://rcs.chph.ras.ru/rcsin.htm [25 February 2006].
35. Chemometry or Chemometrics? (in Macedonian), at the Hemija.net page. http://hemija.net/statii/statija.php?ids = 104 [25 February 2006].
36. What is Chemometrics? (in German), at the page of Chemometrik&QSAR in Würzburg. http://www.pharmazie.uni-wuerzburg.de/AKBaumann/chemometrik.html [25 February 2006].
37. Site map of the PBF.hr at the Food Science and Biotechnology Faculty of Zagreb. http://www.pbf.hr/hr/layout/set/print/content/view/sitemap/2 [25 February 2006].
38. Google search engine. http://www.google.com [6 June 2006].
39. Yahoo search engine. http://www.yahoo.com [6 June 2006].
40. Wikipedia—The Free Encyclopedia. http://www.wikipedia.org/ [6 June 2006].
41. Omniglot—Writing systems & languages of the world. http://www.omniglot.com [6 June 2006].
42. The International Phonetic Association. http://www2.arts.gla.ac.uk/IPA/ipa.html [6 June 2006].
43. Economy Statistics > GDP > Nominal (per capita) by country at Nationmaster.com. http://www.nationmaster.com/graph/eco_gdp_percap-economy-gdp-nominal-per-capita [30 August 2006].
44. GDP per capita for Faroe Islands (2005), Palestine (2003) and Taiwan (2005) at Theodora.com. http://www.theodora.com/wfbcurrent/faroe_islands/faroe_islands_economy.html, http://www.theodora.com/wfbcurrent/west_bank/west_bank_economy.html, http://www.theodora.com/wfbcurrent/taiwan/taiwan_economy.html [1 September 2006]
45. GDP per capita for Cuba (2004 at Photius.com. http://www.photius.com/rankings/economy/gdp_per_capita_2004_1.html [1 September 2006].
46. UNDP Human Development Report 2004. http://hdr.undp.org/reports/global/2004/pdf/hdr04_complete.pdf [25 March 2006].
47. UNDP Human Development Reports: Serbia and Montenegro 2005; http://hdr.undp.org/docs/reports/

Copyright © 2007 John Wiley & Sons, Ltd.

national/YUG_Serbia_and_Montenegro/Serbia_2005_en.pdf, http://hdr.undp.org/docs/reports/national/YUG_Serbia_and_Montenegro/Montenegro_2005_en.pdf [12 April 2006]
48. UNESCO Science Report 2005; http://www.unesco.org/science/science_report2005.pdf [22 April 2006].
49. Tanović L. RTD Status in Bosnia-Herzegovina. http://www.jrc.eu.int/enlargement/events/20041026/presentations/tanovic-geel.pdf [22 April 2006].
50. Human Development Reports 2003: Occupied Palestinian Territories. http://hdr.undp.org/Statistics/data, http://hdr.undp.org/reports/global/2003/indicator/cty_f_PSE.html [23 April 2006].
51. UNCTAD Report 2003: Information and Communication Technology Development Indices. http://www.unctad.org/en/docs/iteipc20031_en.pdf [29 August 2006].
52. ITU Digital Access Index: World's First ICT Ranking 2003. http://www.itu.int/newsarchive/press_releases/2003/30.html [29 August 2006].
53. IANA's Root-Zone Whois Index by TLD Code. http://www.iana.org/cctld/cctld-whois.htm [22 March 2006].
54. Beebe KR, Pell RJ, Sheasholtz MB. *Chemometrics: A Practical Guide*. Wiley: New York, NY, USA, 1998.
55. *Pirouette 3.11*. Infometrix: Woodinville, WA, 2003.
56. *Matlab 6.5*. MathWorks: Natick, MA, 2003.
57. Ethnologue—Languages of the world by SIL International. http://www.ethnologue.com [6 June 2006].
58. Cavalli-Sforza LL. *Genes, Povos e Línguas*. Companhia das Letras: São Paulo, SP, Brazil, 2003.
59. Dyen I, Kruskal JB, Black P. An Indoeuropean classification: a lexicostatistical experiment. *Trans. Amer. Phil. Soc.* 1992; **82**: 1–132.
60. McMahon R. Genes and languages. *Community Genet.* 2004; **7**: 2–13.
61. Cavalli-Sforza LL. Genes, peoples, and languages. *Proc. Natl. Acad. Sci. USA* 1997; **94**: 7719–7724.
62. Greenberg JH. *Indo-European and Its Closest Relatives: The Eurasiatic Language Family*, vol. **1**. Stanford: Stanford, CA, USA, 2000.
63. Austin P, Nancarrow O, Smith G, Stonham J, Woodard R. *O Atlas das Línguas*. Estampa: Lisbon, Portugal, 2001.
64. James HP. Language, Migration, and House Type in Southeast Asia and Beyond. Association for Asian Studies Annual Meeting, March 27–30, 2003, New York, NY, USA. http://www.aasianst.org/absts/2003abst/Border/sessions.htm [25 May 2006].
65. Muturzikin—linguistic maps. http://www.muturzikin.com/en.html [6 June 2006].
66. Ojasoo T. National scientific endeavour from 1981 to 1992. *Sci. Tribune* 1999; (August). http://www.tribunes.com/tribune/art96/ojas1.htm [27 May 2006].
67. Frame JD, Carpenter MP. International research collaboration. *Soc. Stud. Sci.* 1979; **9**: 481–497.
68. Moed HF, De Bruin RE, Nederhof AJ, Tijsse RJW. International scientific cooperation and awareness within the European community: problems and perspectives. *Scientometrics* 1991; **21**: 291–311.
69. Osareh F, Wilson CS. Collaboration in Iranian scientific publications. *Libri* 2003; **52**: 88–98.
70. Moin M, Mahmoudi M, Rezaei N. Scientific output of Iran at the threshold of the 21st century. *Scientometrics* 2005; **62**: 239–248.l.
71. He T, Zhang J, Teng L. Basic research in biochemistry and molecular biology in China: a bibliometric analysis. *Scientometrics* 2005; **62**: 249–259.
72. Kim M-J. Korean science and international collaboration, 1995–2000. *Scientometrics* 2005; **63**: 321–339.
73. Hart PW, Sommerfeld JT. Relationship between growth in gross domestic product (GDP) and the growth in the chemical engineering literature in five different countries. *Scientometrics* 1998; **42**: 299–311.
74. The American Heritage Dictionary of the English Language, 4th edition. http://www.bartleby.com/61/ [1 June 2006].
75. Indo-European Etymological Dictionary at the Leiden University. http://www.indoeuropean.nl/ [1 June 2006].
76. The Tower of Babel—Evolution of Human Language International Project: etymological databases. http://starling.rinet.ru/Intrab.htm [2 June 2006].
77. ARTFL Project at the University of Chicago: Webster's Revised Unabridged Dictionary (1913 + 1828). http://machaut.uchicago.edu/ [11 June 2006].
78. Dictionary.com search service of the LLC. http://dictionary.reference.com/ [9 June 2006].
79. Lavine B, Workman JJ Jr. Chemometrics. *Anal. Chem.* 2004; **76**: 3365–3372.
80. Hemmateenejad B. Chemometrics in Iran. *Chemom. Intell. Lab. Syst.* 2006; **81**: 202–208.
81. Braun T, Glänzel W. International collaboration: will it be keeping alive East European research? *Scientometrics* 1996; **36**: 247–254.
82. Geladi P, Manley M, Lestander T. Scatter plotting in multivariate data analysis. *J. Chemometr.* 2003; **17**: 503–511.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/cem