

HPCI利用研究成果論文のデータ分析 -「京」一般課題、HPCI一般課題- Data Analysis on Research Papers Obtained through the Use of the HPCI System - General Use Categories of K and HPCI -

高度情報科学技術研究機構
 木村 晴行、丸山 順子、平塚 篤

「京」を中核とする革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）によって創出された成果論文に対し、論文の評価指標として広く用いられている相対被引用度（CNCI）を用いた分析を行った。HPCI成果発表データベースに収録されている査読付き論文とWeb of Science（WOS）との照合を定期的に行っており、両者に収録されている「京」を中核とするHPCI計算機資源を用いた成果論文を対象とする（Article, Reviewに限定）。対象とする課題は、公募・選定を経た「京」一般課題、及び「京」以外のHPCI一般課題の内、2012年度～2018年度実施課題（半年課題を除く）とした。課題単位の成果論文のCNCIに基づく成果指標4種；最大CNCI、合計CNCI、平均CNCI、CNCI論文数を設定した。今回の分析では、課題の成果の注目度を反映する最大CNCIの平均値は「京」一般課題では約2.8、HPCI一般課題では約1.7であり、ともに世界平均と同等のパフォーマンスを表す1を十分上回る。成果指標と選定時の評価点群（6種）及び評価に直接関係しない国際共著論文割合等の非評価点群（3種）からなる説明変数（全9種）との間で相関分析、主成分分析、回帰分析を行った。課題の成果の注目度の総量を反映する合計CNCIの回帰分析から、これに寄与する説明変数には「京」一般課題、HPCI一般課題ともCNCI論文数、国際共著論文割合に加え、研究計画の妥当性（評価点群）も含まれることが明らかになり、課題選定の妥当性を支持する分析結果も得られた。

1. はじめに

HPCIは、国内の大学、研究機関の計算機システム、ストレージを高速ネットワークで結ぶことにより、全国のHPCリソースを全国の幅広いHPCユーザー層が効率よく利用できる科学技術計算環境を実現するものである。HPCI計算機資源は特定先端大型研究施設の共用の促進に関する法律に基づく特定高速電子計算機施設（以下、「京」¹）と「京」以外の12の大学、研究機関のスーパーコンピュータから構成される [1]。

一般財団法人 高度情報科学技術研究機構

（RIST）は上記法律に基づく登録施設利用促進機関（以下、登録機関）として、「京」の利用促進業務を実施するとともに、並行して文部科学省委託事業「革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）の運営」に於ける委託事業代表機関の役割を果たして来た [1]。

「京」の利用課題は、公募により選ばれる一般利用枠と文部科学省が定める重点的利用枠（戦略プログラム課題、ポスト「京」研究開発課題、重点化促進枠課題）から構成される。「京」は2012年度下期から共用が開始され、こ

¹ 2019年8月に共用を終了、現在「富岳」の共用開始に向けた準備が行われている。

これらの課題数の合計は約1,010課題(2019年度末時点)である。「京」以外のHPCI計算機資源についても、2012年度下期から利用が開始されており、公募で選ばれる一般利用枠と少数ではあるが重点的利用枠(ポスト「京」研究開発課題)から構成される。これらの課題数の合計は約500課題(2019年度末時点)である。利用者数は全体で年間約2100名(内産業界は450名)に上っている(2018年度の実績)。

「特定高速電子計算機施設の共用の促進に関する基本的な方針」[2]に基づき、RIST神戸センターではHPCI利用研究課題に係るあらゆる成果発表情報を一元的にまとめたHPCI成果発表データベース(以下、成果発表DB)を構築し、2013年5月から公開している[3-4]。このようなハイパフォーマンス・コンピューティング・インフラに於けるPublicationデータベースは米国のXSEDEでは公開されているが(但し論文、会議発表に限られる)[5]、米国のINCITE[6]、欧州のPRACE[7]では公開されていない。

成果発表DBのうち、査読付き論文(英文)に関しては、クラリベイト・アナリティクス社の学術データベースWeb of Science(WOS)との照合を行っている。これにより登録データの信頼性の向上を図るとともに、論文の評価指標として広く用いられている同社InCites Benchmarkingの相対被引用度(CNCI)[8,9]等のデータ収集を行っている。

本報では「京」一般課題、及び「京」以外のHPCI一般課題の成果論文を対象に、CNCIに基づく論文データ分析の結果をまとめる。本報の構成は以下のとおりである；2章では成果発表DBについて、システムの2014年度以降の進展、査読付き論文数の年度発展、WOSとの照合方法・結果について述べる。3章ではCNCIの定義を示すとともに成果指標の設定を行う。4章では対象論文とそのカウント方法を示す。5章ではCNCI値に

係る成果指標の課題の種類単位の統計値を示す。6章では論文データ分析に於ける成果指標に対する説明変数の設定を行う。7章では成果指標と説明変数間の相関分析の結果を示す。8章では成果指標と説明変数の全変数を用いた主成分分析の結果を示す。9章では成果指標の回帰分析の結果を示す。10章、11章で全体をまとめる。

2. HPCI成果発表データベース

2.1. システムの進展

HPCI成果発表データベースは、成果発表検索アプリケーション、成果登録アプリケーションの2つのサブシステムアプリケーションから構成される。

(1) 成果発表検索アプリケーション

成果発表に係る受賞情報を登録できるように拡張した(2015年3月)。これにより利用者は論文や会議発表等と受賞情報を紐づけて閲覧できるようになった。受賞登録は特定の成果発表によらない研究活動の成果に対して贈られた賞もサポートする。

2015年度より開始されたポスト「京」研究開発枠の追加に伴い細分化された利用枠別件数一覧の見づらさを解消するために、大分類毎の利用枠集約表示機能を設けた(2016年7月)(図1参照)。同時に、利用枠別件数表示機能に加え研究分野別件数表示機能を追加した。これにより利用者は2種類の分類別件数を用途に応じて活用できるようになった。

(2) 成果登録アプリケーション

DOI(Digital Object Identifier)データベースからのインポート機能、CSVファイルによる一括入力機能を導入した(2017年5月)。DOIデータベースからのインポート機能は、論文のDOI番号を入力するだけで、DOIデータベースから論文タイトル、著者情報等を自動的にインポートする機能である。CSV一

括入力機能は、入力者が作成したCSV形式データソースを成果発表データベースサーバにアップロードすることで複数の成果情報を一括登録できる機能である。

(3) 収録データ

成果発表DBでは当初より利用者の視点に立ち、HPCI計算機資源を用いた成果のみならず、準備研究などHPCI以外の計算機資源を用いた成果や、理論研究など計算機を使用しない成果の発表情報も登録できる。図2に、(a) 全発表カテゴリ、(b) 査読付き論文の登録データ件数(累積値)の年度推移をHPCI計算機資源、HPCI以外の計算機資源、計算機資源使用無に分類して示す(HPCI計算機資源、HPCI以外の計算機資源の各件数には両者利用による重複を一部含む)。2019年度までの重複を排除した全登録総数は12,442件(うちHPCI計算機資源は8,430件)、うち査読付き論文は3,250件(うちHPCI計算機資源は2,028件)に達する(2020/6/11現

在)。査読付き論文は当初はHPCI以外の計算機資源の論文数がHPCI計算機資源の論文数を圧倒的に上回っていたが、2016年度にHPCI計算機資源の論文数が逆転していることが分かる。

なお、査読付き論文(英文)については過去に遡りDOI番号を再調査しDOI番号に紐づけた。これにより存在していた幾つかの重複データを排除することが出来た(2.3参照)。

2.2. 査読付き論文数の年度発展

図3に、2020年5月31日現在において成果発表DBに登録されている、HPCI計算機資源を用いて得られた査読付き論文の件数を、論文発表年度毎に示す。色分けは「課題の種類」毎の件数である(各論文件数には課題の種類間で一部重複がある)。HPCI戦略プログラムは2015年度末で終了し、ポスト「京」研究開発枠重点課題は2015年10月1日から、ポスト「京」研究開発枠萌芽的課題は2016年8月1日からスタートしている。

	課題の種類(複数の課題が関与する成果発表は各「課題の種類」の件数としてカウントされます) 表示オプション						合計	正味の成果発表件数
	「京」一般利用	「富岳」・「京」以外のHPCI一般利用	「富岳」成果創出加速プログラム	ポスト「京」研究開発枠重点課題	ポスト「京」研究開発枠萌芽的課題	HPCI戦略プログラム		
	Expand	Expand	Expand	Expand	Expand	Expand		
査読付き論文	626	664	0	557	99	548	2494	2035
査読なし論文	43	30	0	39	3	88	203	187
国際会議・シンポジウム	599	383	0	559	29	977	2547	2128
国内学会・シンポジウム	505	276	0	682	43	916	2422	2079
研究会等	224	117	0	229	5	346	921	813
一般向け講演会等	94	49	0	141	0	178	462	413
新聞/TV/雑誌/WEB配信等	104	34	0	129	0	424	691	644
書籍	16	12	0	42	2	9	81	72
プログラム・DB公開	6	0	0	3	0	4	13	11
特許出願	3	2	0	0	0	5	10	10
特許取得	7	4	0	3	0	6	20	17
合計	2227	1571	0	2384	181	3501	9864	8409

図1 成果発表DBトップ画面(2020年6月11日現在);縦方向に発表のカテゴリ、横方向に大分類毎に集約した利用枠を示す。各セルの数字は発表件数であり、[+] Expandをクリックすると大分類が展開され個別の件数が表示される。各セルの数字をクリックすると成果発表リストが表示され、各リストの行をクリックすると個々の成果発表情報の詳細が表示される[3]。

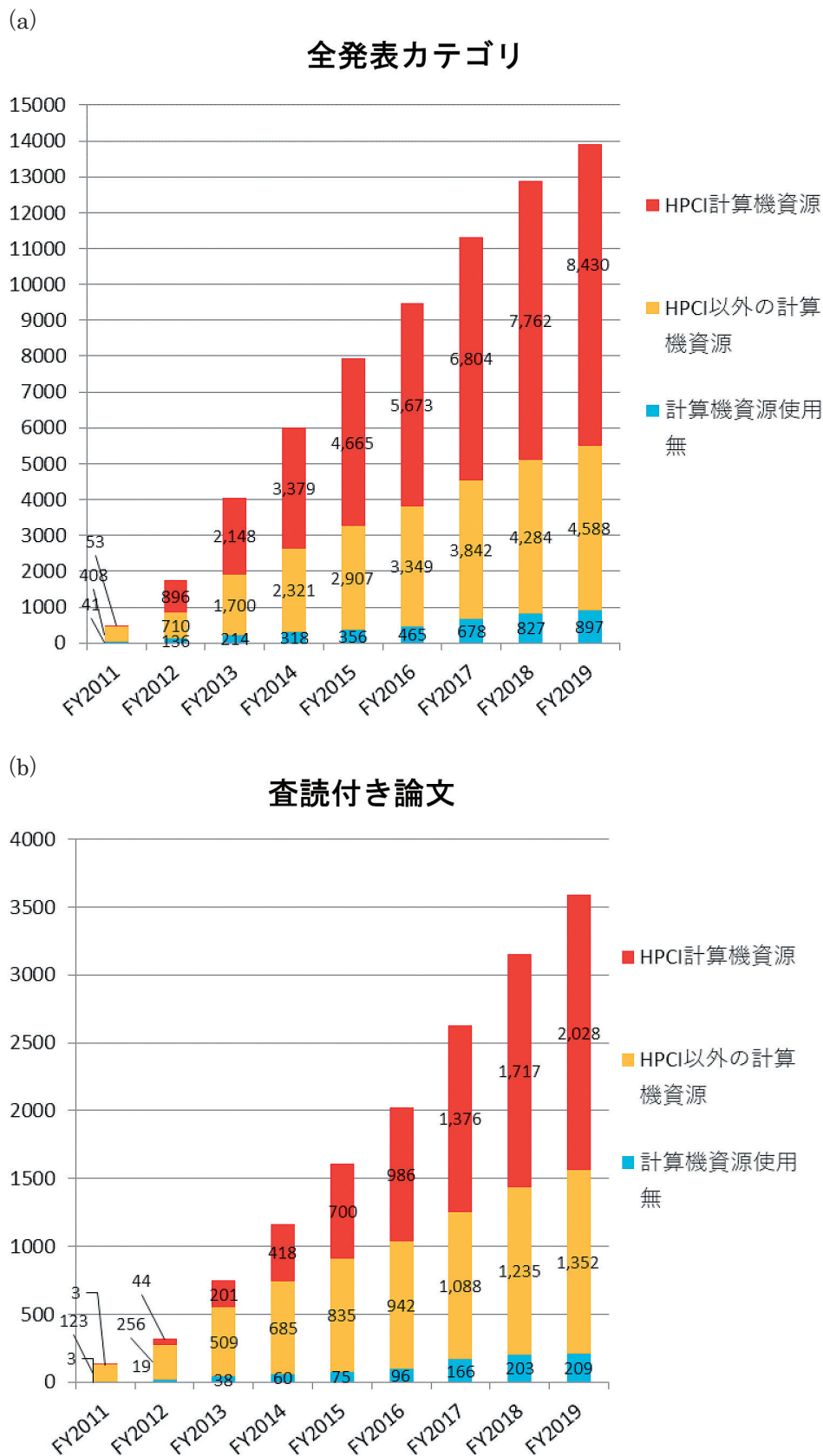


図2 成果発表DBの登録データ件数(累積値)の年度推移
(a) 全発表カテゴリ、(b) 査読付き論文

2015年度まではHPCI戦略プログラム関係の発表件数が最も多い。2016年度は「京」一般利用関係の発表件数が最も多く、2017年度以降はポスト「京」研究開発枠重点課題関係の発表件数が最も多くなっている。これは、「京」からポスト「京」に係る研究開発へと成果発表対象が移行していることを表している。一方ポスト「京」研究開発枠萌芽的課題関係の発表件数が少ない理由としては、挑戦的な取り組みであることから短期間での研究成果の論文化が難しいことが考えられる。

2.3. Web of Scienceとの照合

2017年度より、成果発表DBとWOS収録論文との照合を行い、WOSに収録されているHPCI利用研究課題（以下、「HPCI課題」という。）に関連する論文で成果発表DBに未登

録の論文を抽出する取り組みを行っている。WOSに収録されているこれらの論文の抽出方法については、WOSの「助成金提供機関と助成金登録番号」および「助成金情報」に、課題番号またはHPCI課題関係であると特定できるキーワードが含まれる論文を抽出している（この論文データセットをWOS (HPCI) と呼ぶ）。また、その結果と成果発表DBとを照合する情報として、タイトルおよびDOI番号を用いている。

表1に、WOS (HPCI) (2,967件) と成果発表DB (査読付き論文、査読なし論文) との照合結果を使用計算機資源別（「京」を含むHPCI計算機資源、HPCI以外の計算機資源、計算機資源使用無の3分類）に示す（2020/5/24現在）。WOS (HPCI) 抽出結果のうち1,821件が成果発表DB査読付き論文に登録済

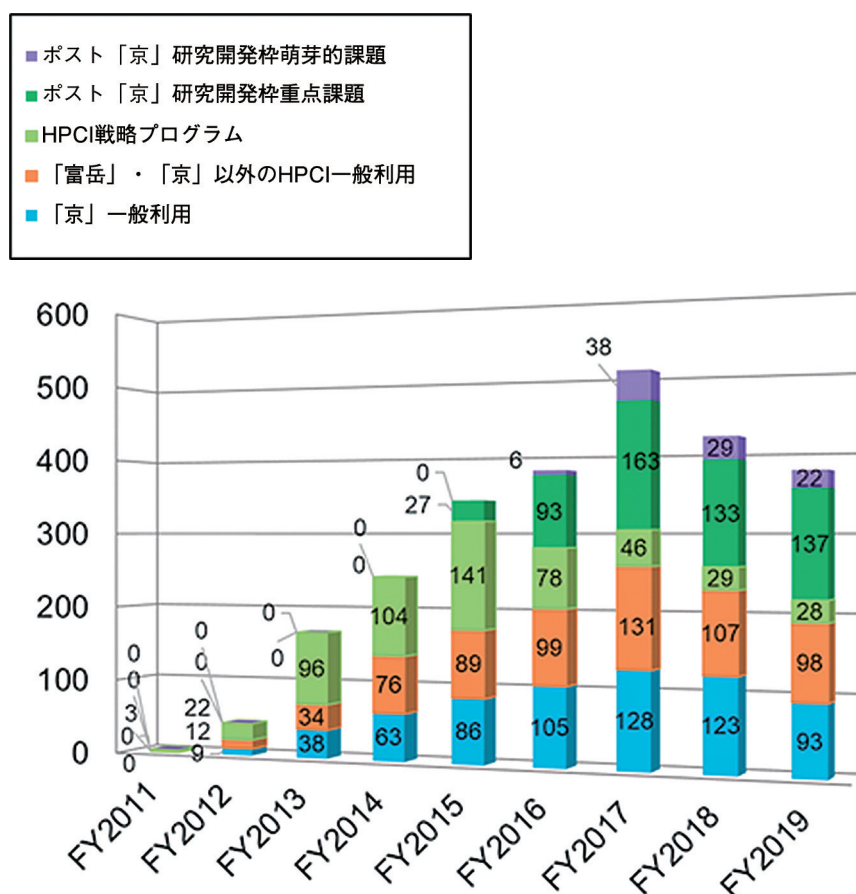


図3 成果発表DBの査読付き論文発表件数
(論文発表年度毎)

み、5件が同査読なし論文に登録済み、1,131件が未登録である（なお、表1の内訳には示していないが成果発表DBの「京」調整高度化枠利用課題²にのみ登録されている論文が10件ある）。査読付き論文にはそれぞれ課題番号が記載された論文数（内訳）が示されている。その記載割合はHPCI計算機資源では87%と高い値を示すが、HPCI以外の計算機資源、計算機資源使用無ではそれぞれ32%、21%である。成果発表DB未登録の論文に於ける課題番号記載率も5%と低い。成果発表DB未登録論文の使用計算機資源は登録されるまで未知であるが、それぞれの課題番号記載割合が成果発表DB登録済論文と同程度であると仮定することにより、使用計算機資源の割合を推定出来る。これによれば、成果発表DB未登録の論文はHPCI以外の計算機資源、計算機資源使用無の論文が大半を占める（9割以上）と推定される。

一方、成果発表DBに登録されている査読

付き論文でWOSに収録されているもののWOS（HPCI）に含まれない（助成金情報等にHPCIに関するキーワードがない又は明確でない）論文が少なからず（709件）存在する（この論文データセットをWOS（others）と呼ぶ）。表2に成果発表DBの査読付き論文登録数とそのうちWOSに収録されている論文数の内訳を示す。成果発表DB登録論文数全体2,948件（うちHPCI計算機資源は1,779件）に対し2,530件（うちHPCI計算機資源は1,509件）がWOSに収録されている（2020/5/24現在）。

3. 相対被引用度と成果指標の設定

3.1. 相対被引用度

我が国を含み世界の研究評価基準で使われるクラリベイト・アナリティクス社の研究分析データベースInCites Benchmarking [8]に於いて、相対被引用度（CNCI）は下記式によって定義されている [9]。

表1 WOSから抽出したHPCI課題関係のキーワードが含まれる論文データセット（WOS（HPCI））と成果発表DBの照合結果

	全体	WOS(HPCI)						
		成果発表DB (査読付き) に有	同左 課題番号有	課題番号有 の割合	成果発表DB (査読なし)に 有	成果発表DB に無	同左 課題番号有	課題番号有 の割合
全体	2,967	1,821	1,283	0.70	5	1,131	60	0.05
HPCI計算機資源		1,280	1,118	0.87	2			
HPCI以外の計算機資源		471	150	0.32	3			
計算機資源の使用無		70	15	0.21	0			

表2 成果発表DBとWOS収録論文

査読付き論文	成果発表DB 登録論文数	うちWOS収録論文数		
		合計	WOS(HPCI)	WOS(others)
全体	2,948	2,530	1,821	709
HPCI計算機資源	1,779	1,509	1,280	229
HPCI以外の計算機資源	964	866	471	395
計算機資源の使用無	205	155	70	85

² 「京」の運用機関（理化学研究所）による運転調整や高度化研究に利用可能な計算資源を用いた課題

$CNCI = \text{Times Cited} / \text{Category Expected Citations}$
 ここで、Times Citedは被引用数、Category Expected Citationsは研究領域、出版年、およびドキュメントタイプが考慮された期待引用率である。

CNCIの特徴は、論文の出版年が異なっても比較が可能な評価指標であり、 $CNCI = 1$ の論文は世界平均と同等のパフォーマンスを表し、1を超える値は平均を上回るとみなされる[9]。なお、本報に於けるCNCIデータはWeb of Scienceの研究領域分類に基づいている。

3.2. 成果指標の設定

HPCI利用研究課題に対しCNCIをベースに以下の4つの成果指標を設定する；

最大CNCI (Z)：課題単位の成果論文の

CNCIの最大値→成果の注目度を反映

合計CNCI (A)：課題単位の成果論文の

CNCIの合計値→成果の注目度の総量を反映

平均CNCI (B)：課題単位の成果論文の CNCIの平均値→成果の平均的な注目度を反映

CNCI論文数 (C)：課題単位のCNCIが定義された成果論文の数→成果の量を反映

この4つの指標を後述(9章)の回帰分析に於ける目的変数とする。

4. 対象論文とカウント方法

文部科学省 科学技術・学術政策研究所(NISTEP)の論文分析資料[10]に合わせて、論文種別をArticleとReviewに限定する。

複数課題が関与する論文に対して各課題の論文として1個ずつカウントする(課題単位の整数カウント)。

表3に対象とする「京」一般課題³、HPCI一般課題⁴の2018年度までの全実施課題数(一部を除く)、CNCI論文有の課題数、全実施課題数に対するその割合、CNCI論文数の総和、CNCI論文数の平均値、を示す。後述の相関分析、主成分分析、回帰分析では突出した成

表3 課題の種類単位の実施課題数、CNCI論文数に関する各種数値データ (2020/4/15時点)

課題の種類	実施課題数 (A)	CNCI論文有の課題数(B)	CNCI論文有の課題割合(B/A)	CNCI論文数の総和 ^{*1} (C)	CNCI論文数の平均値(C/B)
「京」一般課題	156 (154)	119 (117)	76.3 % (76.0 %)	333 (325)	2.80 (2.78)
HPCI一般課題	345 (341)	218 (214)	63.2% (62.8 %)	603 (582)	2.77 (2.72)

集計範囲；「京」一般課題(2012年度実施課題 (hp12) - 2018年度A期実施課題 (hp18A) (半年課題であるhp13、hp17Aを除く))、HPCI一般課題 (hp12-hp18 (半年課題であるhp13を除く))、いずれもHPCI計算機資源利用成果、数値行の括弧内は外れ値(2課題(「京」一般課題)、4課題(HPCI一般課題))を除外した数字

*1 課題毎のCNCI論文数の総和

³ 「京」一般利用枠全体は一般課題、一般トライアル・ユース、競争的資金等獲得課題、若手人材育成課題、産業利用課題(実証利用、トライアル・ユース、個別利用、ASP事業実証利用)、共用法第12条調査研究から構成される。「京」一般課題とはこのうちの一般課題を意味する。

⁴ HPCI一般利用枠全体は一般課題、若手人材育成課題、産業利用課題(実証利用、トライアル・ユース、個別利用)、HPCI共用ストレージ利用課題から構成される。HPCI一般課題とはこのうちの一般課題を意味する。

果指標を有する課題を除外している。「京」一般課題では2課題（合計CNCI \geq 50または最大CNCI \geq 25の課題）、HPCI一般課題では4課題（合計CNCI \geq 20または最大CNCI \geq 10の課題）である。これら外れ値除外後の数値を括弧内に示している。CNCI論文数の平均値は「京」一般課題とHPCI一般課題で同程度（約2.8）である。

図4にCNCI論文数の度数分布（縦軸は課題数）を示す。「京」一般課題とHPCI一般課題でCNCI論文数の平均値は同程度であるが、「京」一般課題の方がピーク近傍では緩やかな分布になっている。HPCI一般課題では少数であるが極めて多数の論文を産生している課題が存在する。

る課題が存在する。

5. CNCI値に係る成果指標の統計値

表4に課題の種類単位の最大CNCIの総和、同合計CNCIの総和、それぞれをCNCI論文有の課題数で割った平均値、課題の種類単位のCNCI論文の総数（重複無）、同CNCI値の総和（重複無）、同全体平均値を示す。ここで、平均CNCIについては平均CNCIの平均値ではなく全体平均値を用いている。

成果の注目度を反映する最大CNCIの平均値は「京」一般課題が約2.8、HPCI一般課題は約1.7である。CNCI=1が世界平均と同等のパフォーマンスを表し、いずれもこれを

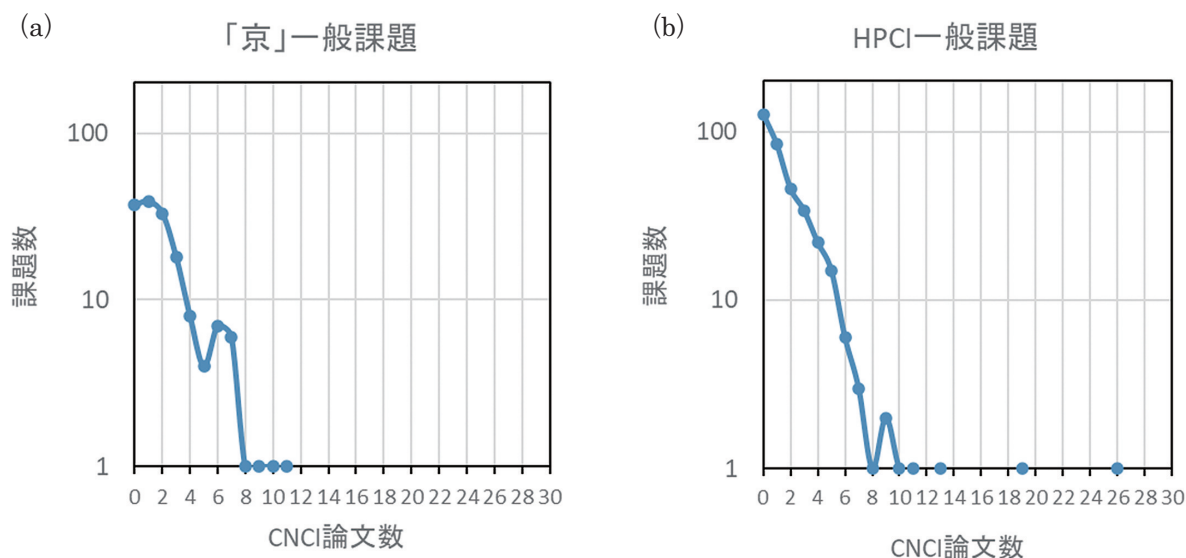


図4 CNCI論文数の度数（課題数）分布（対数目盛）
 (a) 「京」一般課題全体（156課題）、(b) HPCI一般課題全体（345課題）

表4 課題の種類単位のCNCI値に係る成果指標に関する各種数値データ（2020/4/15時点）

課題の種類	最大CNCIの総和 ^{*1} (D)	合計CNCIの総和 ^{*2} (E)	最大CNCIの平均値(D/B)	合計CNCIの平均値(E/B)	CNCI論文総数 ^{*3} (F)	CNCI総和 ^{*4} (G)	CNCIの全体平均値(G/F)
「京」一般課題	331.3 (260.8)	572.6 (469.4)	2.78 (2.23)	4.81 (4.01)	235 (227)	404.5 (301.3)	1.72 (1.33)
HPCI一般課題	372.1 (333.8)	639.7 (560.0)	1.71 (1.56)	2.93 (2.62)	400 (391)	442.6 (410.5)	1.11 (1.05)

集計範囲は表3と同じ。タイトル行の括弧内アルファベット記号は表3から継続。数値行の括弧内は外れ値（2課題（「京」一般課題）、4課題（HPCI一般課題））を除外した数字

*1 課題毎の最大CNCIの総和、*2 課題毎の合計CNCIの総和、

*3 課題の種類単位のCNCI論文の総数（重複無）、*4 課題の種類単位のCNCI値の総和（重複無）

十分上回っている。

成果の注目度の総量を反映する合計CNCIの平均値は「京」一般課題が約4.8、HPCI一般課題は約2.9である。このことは「京」一般課題では世界平均と同等のパフォーマンスを表すCNCI=1の論文を平均約5編、HPCI一般課題では平均約3編公開したと等価であることを意味する。

成果の平均的な注目度を反映するCNCIの全体平均値は「京」一般課題が約1.7、HPCI一般課題は約1.1であり、ともに世界平均と同等のパフォーマンスを表す1を上回っている。

図5に3成果指標の度数分布を示す（縦軸は課題数）。各成果指標の平均値、外れ値を除外した平均値、及び中央値の位置も、それぞれ赤線、オレンジ線、及び緑線で示されている。「京」一般課題、HPCI一般課題とも、

これら3成果指標の度数分布はやや平坦なべき乗分布をしていることが分かる。特に、合計CNCIの場合は「京」一般課題、HPCI一般課題ともすそ野が高い値にまで伸びている。なお、「京」一般課題については、外れ値2課題のうち最大CNCI、合計CNCIで各2課題、平均CNCIで1課題が横軸の上限を超えるため表示されていない。HPCI一般課題については、最大CNCIで外れ値4課題のうち2課題が表示されていない。

図6は図5の度数分布に対応する課題数の規格化累積値分布である。これにより、「京」一般課題に於いて各成果指標が平均値の課題は全体の上位3割程度に位置することが分かる。外れ値を除外した平均値の課題は最大CNCI、平均CNCIでは上位4割程度に、合計CNCIでは全体の上位1/3程度にそれぞれ位置する。HPCI一般課題に於いても同様に各

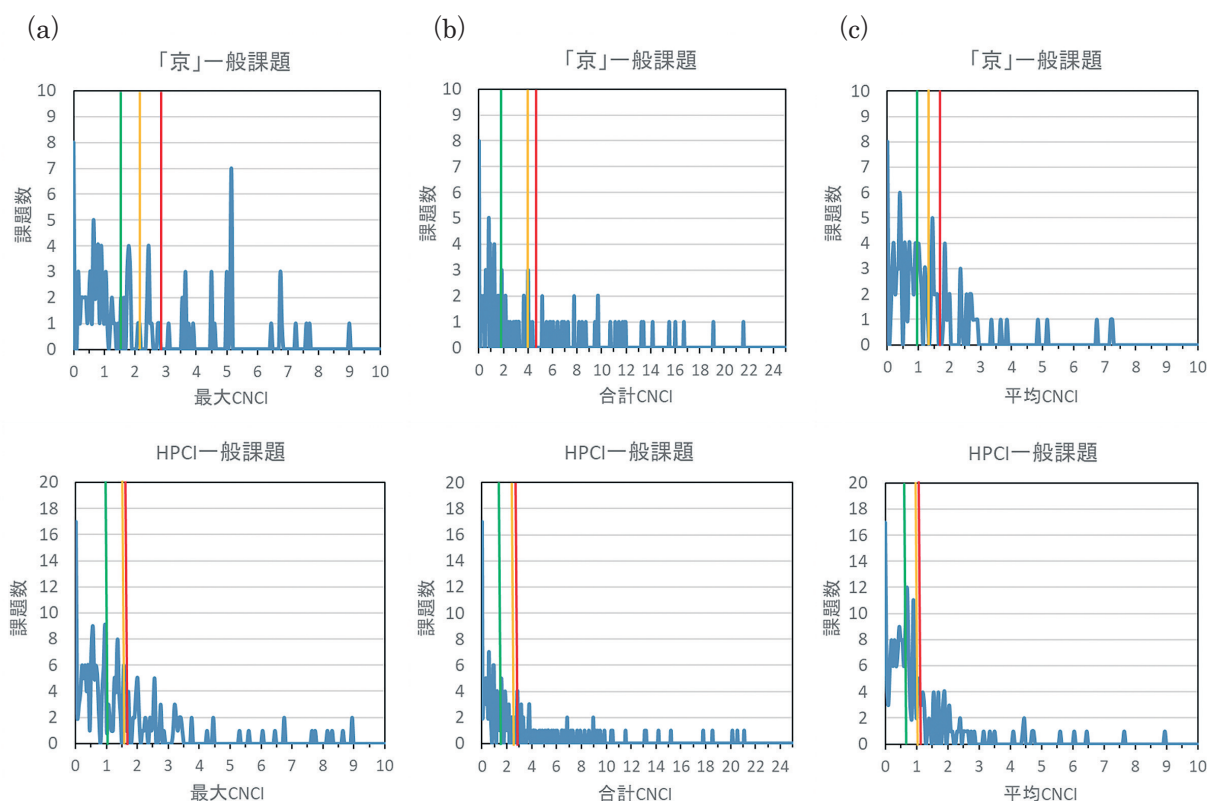


図5 CNCI値に係る成果指標の度数（課題数）分布、(a) 最大CNCI、(b) 合計CNCI、(c) 平均CNCI。上段は「京」一般課題（CNCI論文有の全課題）、下段はHPCI一般課題（CNCI論文有の全課題）、赤線は各全体の中での平均値の位置、オレンジ線は外れ値を除外した平均値の位置、緑線は各全体の中での中央値の位置を示す。合計CNCIの横軸の範囲は他の2.5倍としている。

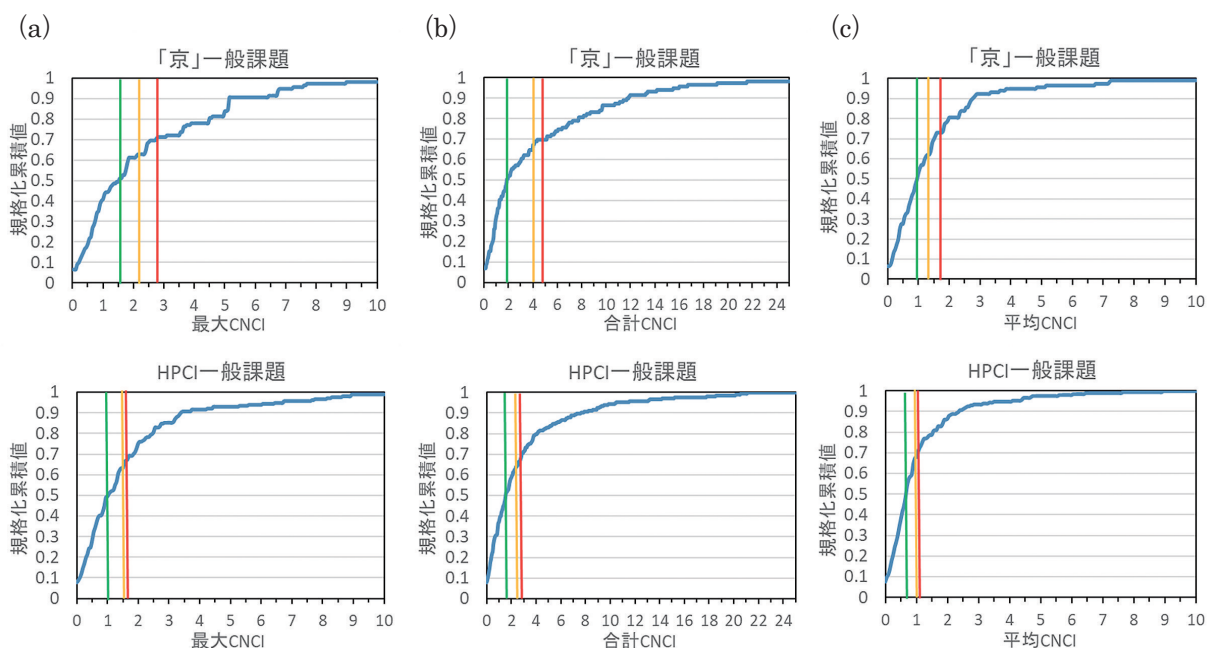


図6 図5に対応する課題数の規格化累積値分布、(a) 最大CNCI、(b) 合計CNCI、(c) 平均CNCI

成果指標が平均値の課題は全体の上位3割程度に位置する。外れ値を除外した平均値については最大CNCI、合計CNCIでは全体の上位1/3程度に位置する。平均CNCIでは両者の位置は殆ど変わらない。

平均値と中央値との比較を行うと、「京」一般課題、HPCI一般課題とも最大CNCIと平均CNCIに比べて、合計CNCIの場合(特に、「京」一般課題)は平均値と中央値の差が相対的に開く傾向にあることが分かった。これは上述の合計CNCIの場合は「京」一般課題、HPCI一般課題ともすそ野が高い値にまで伸びていることに対応する。

6. 説明変数

説明変数は課題選定時の評価点からなる評価点群とそれ以外の非評価点群から構成される。評価点群は個別評価点(D)及び5つの個別要素点(科学的・社会的意義(E)、「京」(またはHPCI)の必要性(F)、ソフトウェアの準備状況(G)、研究計画の妥当性(H)、研究体制の妥当性(I))の計6個の変数で構成される。なお、個別評価点は個別要素点か

ら算出される。

非評価点群は課題参加者数(J)、国際共著論文割合(K)、重点的利用枠との共著論文割合(L)の3つの変数で構成される。

7. 成果指標と説明変数間の相関分析

成果指標(Z, A, B, C)と説明変数(評価点群; D~I、非評価点群; J~L)間の相関関係を「京」一般課題の場合は図7に、HPCI一般課題の場合は図8に示す(それぞれCNCI論文有の課題に限定)。線の太さで各変数間の相関係数の大きさを示す。白抜きの線は相関係数がマイナスであることを示す。

特徴的なことを以下に記す; 成果指標(特にZ, A, B)間の相関は「京」一般課題、HPCI一般課題ともに強い。国際共著論文割合(K)と成果指標(Z, A, B, C)との相関は「京」一般課題の方が強い。また、個別評価点(D)と成果指標(Z, A, B, C)、国際共著論文割合(K)及び重点的利用枠との共著論文割合(L)との相関は「京」一般課題では一定程度ある。これらの相関はHPCI一般課題の場合、「京」一般課題に比べて明ら

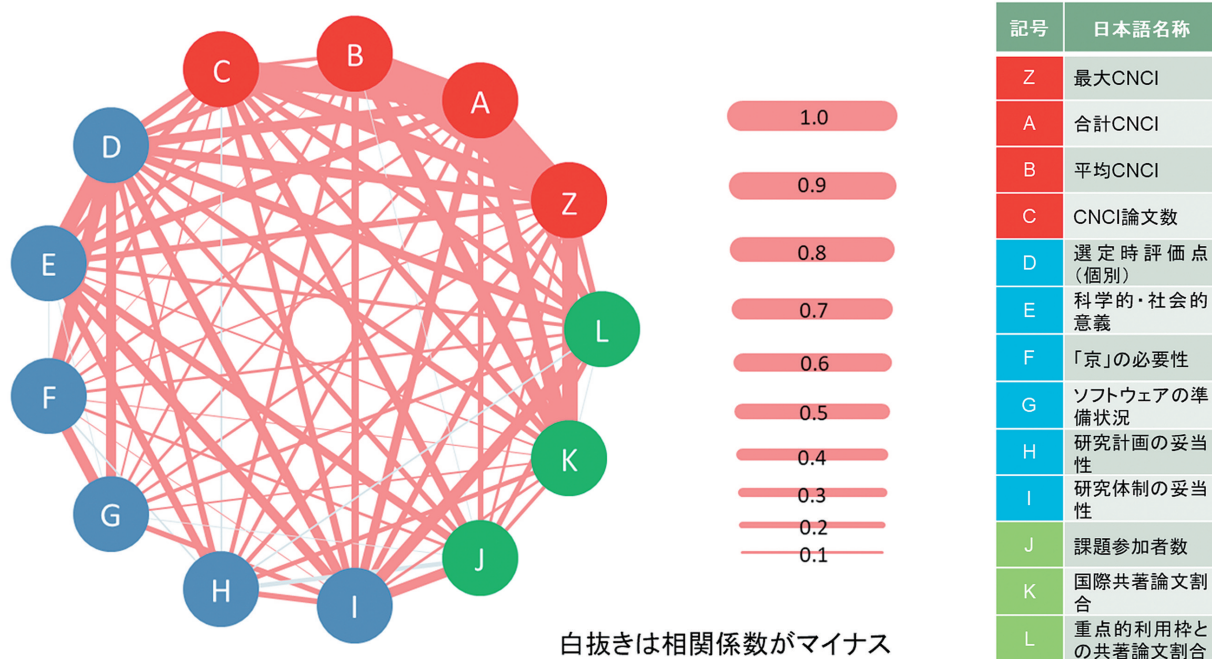


図7 成果指標と説明変数間の相関係数（「京」一般課題（CNCI論文有117課題、外れ値（2課題）除外））

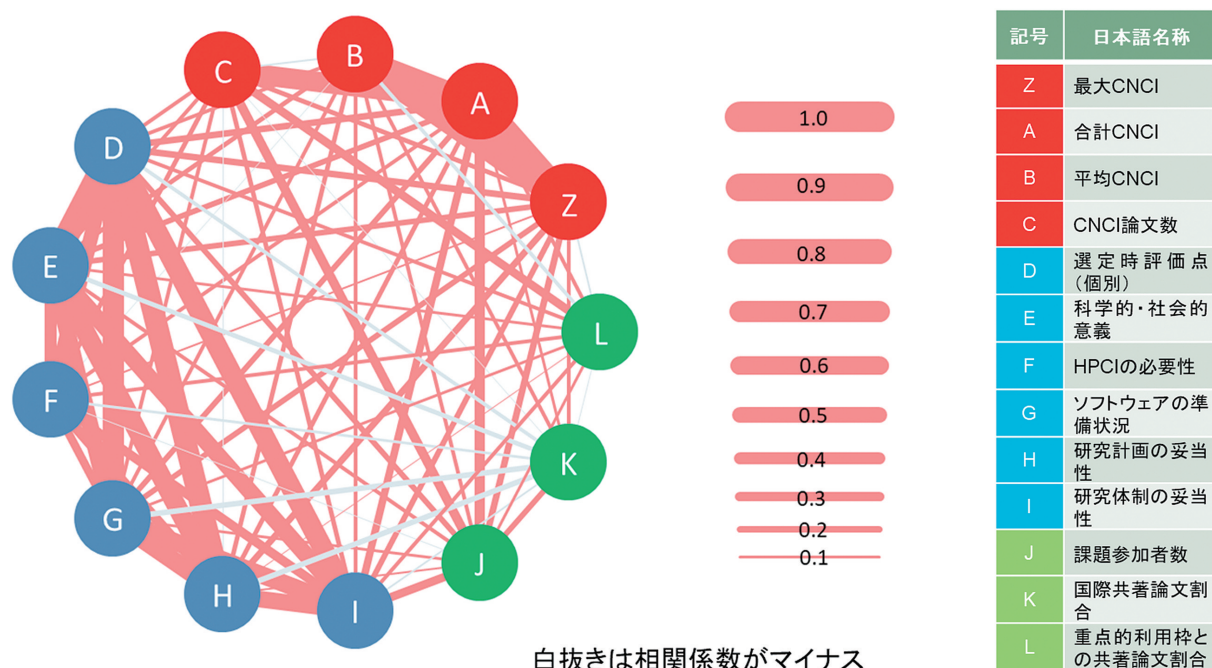


図8 成果指標と説明変数間の相関係数（HPCI一般課題（CNCI論文有214課題、外れ値（4課題）除外））

かに弱い。一方、課題参加者数（J）については、個別評価点（D）、成果指標（C）との相関はHPCI一般課題と「京」一般課題は同程度であるが、成果指標（Z, A, B）との相関はHPCI一般課題の方が「京」一般課題より強い。

評価点群（D～I）内の相関はHPCI一般課題では極めて強く、「京」一般課題では弱い。

なお、CNCIのデータは3ヵ月ごとに更新されている。我々は2019年度4月にCNCIデータ（2019/3/29付けデータ）の取得を開始し、その後3ヵ月ごとにデータ更新を行って

る。データ取得開始後ほぼ1年を経過して変動は殆ど見られなくなっている。本報では現時点の最新のデータ(2020/3/26付けデータ)を用いている。

8. 全変数の主成分分析

「京」一般課題及びHPCI一般課題について、成果指標、評価点群、非評価点群の関係を総合的に評価するために、これら全変数を対象として主成分分析(PCA; Principal component analysis)を行った[11]。PCAは多変量解析の基本的な手法である。多次元データの情報をなるべく損なわないように低次元に集約する手法で、データの相関行列について固有値問題を解くことにより実施した。

8.1. 「京」一般課題の場合

「京」一般課題(CNCI論文有の課題に限定)の結果を図9(a)～(e)に示す。各図の主たるポイントは以下のとおりである；

(a) 第1主成分(PC1)に全体の30%の情報が縮約され、第2主成分(PC2)までで全体の44%の情報、第3主成分(PC3)までで全体の55%の情報、第6主成分(PC6)までで全体の79%の情報を占める。

(b) 成果指標(Z, A, B)と国際共著論文割合(K)はPC1とPC2に同相で寄与する。

(c) 全変数は以下の4グループに分類できる；

- ・成果指標(Z, A, B)と国際共著論文割合(K)
- ・成果指標(C)
- ・研究計画の妥当性(H)
- ・(H)以外の評価点群(D, E, F, G, I)、課題参加者数(J)及び重点的利用枠との共著論文割合(L)

(d) 成果指標(Z, A, B)、国際共著論文割合(K)の固有ベクトルが一群となり主としてPC1の方向に向いている。研究計画の妥当性(H)以外の評価点群(D, E, F, G, I)、課題参加者数(J)及び重点的利用枠と

の共著論文割合(L)の固有ベクトルが一群となり主としてPC2の方向に向いている。成果指標(C)と研究計画の妥当性(H)の固有ベクトルはその中間部に位置し、ほぼPC1に沿っている。成果指標の低い領域(薄緑、黄色の領域)に存在する課題のスコアプロットはHPCI一般課題の場合(図10(d))に比べると疎である。

(e) PC1とPC3の因子負荷量プロット上に於いて成果指標(Z, A, B, C)と評価点群(D, E, I)、国際共著論文割合(K)、重点的利用枠との共著論文割合(L)が接近している。

上記のことから、「京」一般課題の場合は成果指標(Z, A, B, C)と国際共著論文割合(K)、研究計画の妥当性(H)との間に比較的強い相関がある。また、成果指標(Z, A, B, C)と他の評価点群(D, E, I)、重点的利用枠との共著論文割合(L)との相関も認められる。

8.2. HPCI一般課題の場合

HPCI一般課題(CNCI論文有の課題に限定)の結果を図10(a)～(e)に示す。各図の主たるポイントは以下のとおりである；

(a) PC1に全体の32%の情報が縮約され、PC2までで全体の51%の情報、PC3までで全体の63%の情報、PC6までで全体の84%の情報を占める。

(b) 成果指標(Z, A, B, C)と課題参加者数(J)はPC1とPC2に同相で寄与する。

(c) 全変数は以下の4グループに分類できる；

- ・成果指標(Z, A, B)
- ・成果指標(C)と課題参加者数(J)
- ・国際共著論文割合(K)
- ・評価点群(D, E, F, G, H, I)と重点的利用枠との共著論文割合(L)

(d) 成果指標(Z, A, B, C)及び課題参加者数(J)の固有ベクトルは一群となりPC

(a)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Standard deviation	1.9862	1.3225	1.1868	1.1741	0.9884	0.9057	0.8418	0.7915	0.7361	0.6470	0.5262	0.2937	0.2504
Proportion of Variance	0.3035	0.1345	0.1084	0.1061	0.0752	0.0631	0.0545	0.0482	0.0417	0.0322	0.0213	0.0066	0.0048
Cumulative Proportion	0.3035	0.4380	0.5464	0.6524	0.7276	0.7907	0.8452	0.8934	0.9350	0.9672	0.9885	0.9952	1.0000

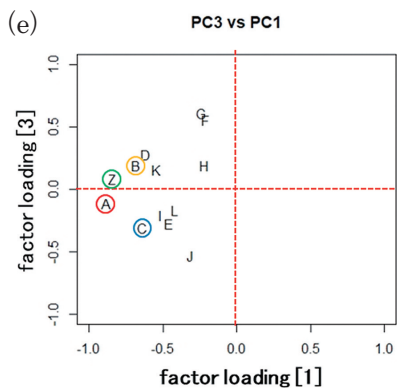
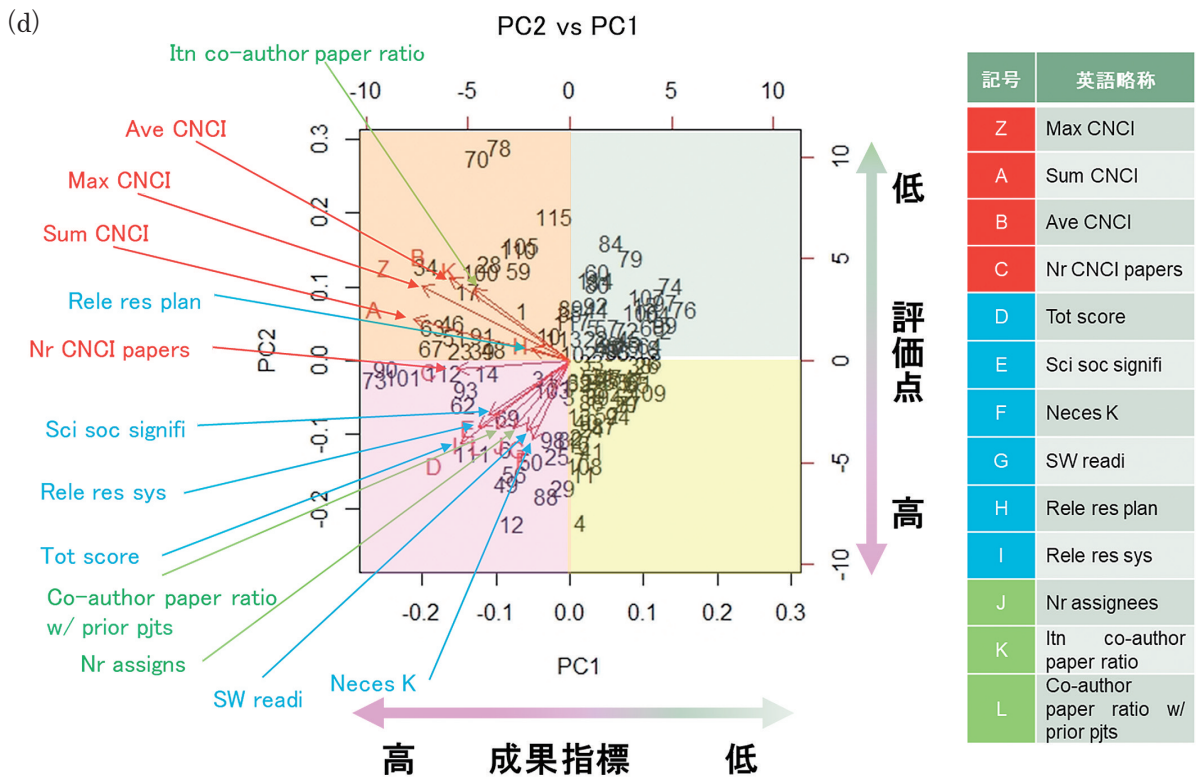
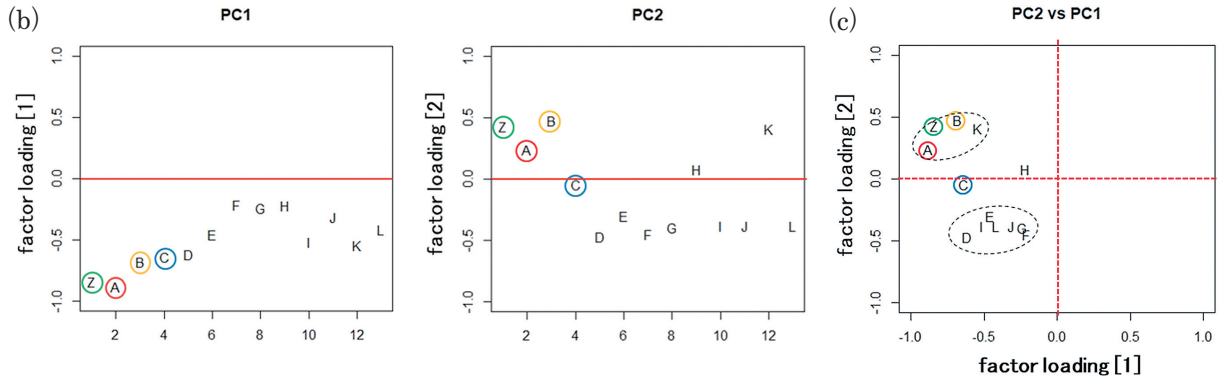


図9 (a) 各主成分の標準偏差、寄与率、累積寄与率、
 (b) PC1、PC2に関する因子負荷量プロット (1次元)、
 (c) 同因子負荷量プロット (2次元)、
 (d) 同Biplot図 (外れ値を除くCNCI論文有課題のスコアプロットと全変数の固有ベクトルを示す)、
 (e) PC1、PC3に関する因子負荷量プロット (2次元)
 (b)、(c)、(e) に於いて、成果指標は色付けた○で囲んでいる。「京」一般課題 (CNCI論文有117課題、外れ値 (2課題) 除外)

(a)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Standard deviation	2.0549	1.5559	1.2498	1.0422	0.9243	0.8873	0.7717	0.6741	0.6232	0.5849	0.3844	0.3100	0.2049
Proportion of Variance	0.3248	0.1862	0.1202	0.0836	0.0657	0.0606	0.0458	0.0350	0.0299	0.0263	0.0114	0.0074	0.0032
Cumulative Proportion	0.3248	0.5111	0.6312	0.7148	0.7805	0.8411	0.8869	0.9218	0.9517	0.9780	0.9894	0.9968	1.0000

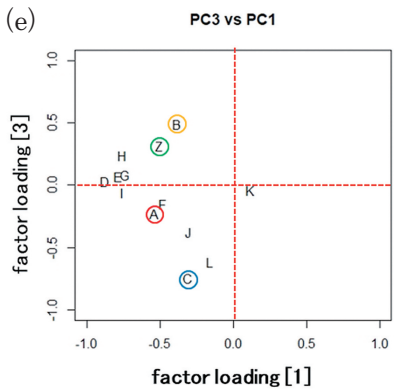
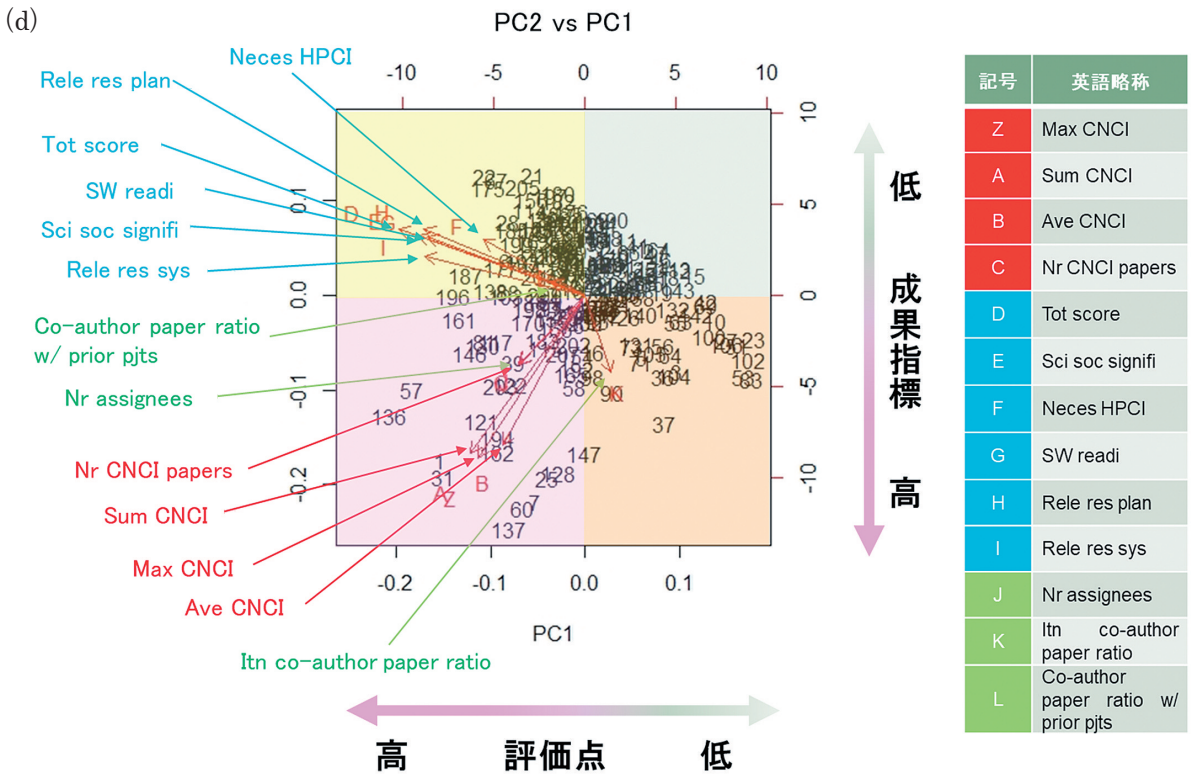
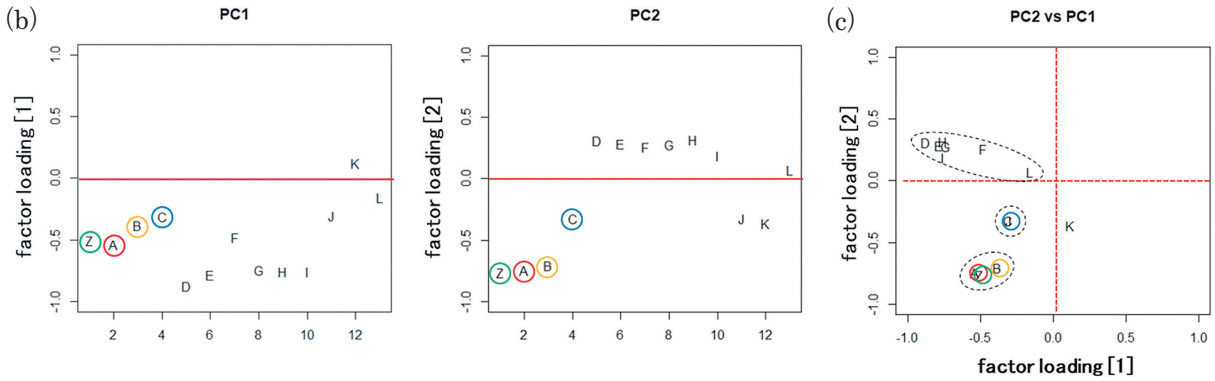


図10 (a) 各主成分の標準偏差、寄与率、累積寄与率、
 (b) PC1、PC2に関する因子負荷量プロット (1次元)、
 (c) 同因子負荷量プロット (2次元)、
 (d) 同Biplot図 (外れ値を除くCNCI論文有課題のスコアプロットと全変数の固有ベクトルを示す)、
 (e) PC1、PC3に関する因子負荷量プロット (2次元)
 (b)、(c)、(e) に於いて、成果指標は色付けした○で囲んでいる。HPCI一般課題 (CNCI論文有214課題、外れ値 (4課題) 除外)

1とPC2の中間方向を向いている。国際共著論文割合（K）の固有ベクトルは成果指標の固有ベクトルとやや離れて主としてPC2の方向に向いている。評価点群（D, E, F, G, H, I）と重点的利用枠との共著論文割合（L）の固有ベクトルは一群となり成果指標（Z, A, B, C）の固有ベクトルとほぼ直角方向を向いている。成果指標の低い領域（薄緑、黄色の領域）に存在する課題のスコアプロットは「京」一般課題の場合（図9（d））に比べて密である。

(e) PC1とPC3の因子負荷量プロット上に於いて成果指標（Z, A, B）と評価点群（D, E, F, G, H, I）が接近し、また、成果指標（C）と重点的利用枠との共著論文割合（L）が接近している。

上記のことから、HPCI一般課題の場合は成果指標（Z, A, B, C）と課題参加者数（J）との間に比較的強い相関がある。また、成果指標（Z, A, B）と国際共著論文割合（K）との相関、成果指標（Z, A, B）と評価点群（D, E, F, G, H, I）との相関、成果指標（C）と重点的利用枠との共著論文割合（L）との相関もそれぞれ認められる。

9. 成果指標の回帰分析

目的変数である各成果指標に対して、これとの相関が比較的強い説明変数の組合わせ（他の成果指標も含む）を主成分分析（PCA）によって抽出し、それらの候補を用いて、決定係数の最大化、標準誤差の最小化、有意確率（P値）の最小化などの観点から最適と見なしうる重回帰分析結果を求める手法を取る。なお、本章で示す回帰分析結果について、より精度の高い分析手法である部分的最小二乗回帰（PLSR）[12]を用いた検証を進めており、これまでのところ両者がほぼ矛盾の無い結果を示すことを確認している。

9.1. 「京」一般課題の場合

「京」一般課題（CNCI論文有の課題に限定）の結果を図11（a）～（d）に示す。目的変数は（a）が最大CNCI、（b）が合計CNCI、（c）が平均CNCI、（d）がCNCI論文数である。（a）、（b）、（c）いずれもこれにCNCI論文数、評価点群（6種）、非評価点群（3種）を加えた全11種の変数によるPCAを行っている。（d）では論文発表とCNCI値の因果関係を踏まえて、CNCI値に係る成果指標は説明変数に用いず、CNCI論文数に評価点群（6種）、非評価点群（3種）を加えた全10種の変数によるPCAを行っている。得られた因子負荷量プロット（2次元；PC2 vs PC1、PC3 vs PC1）を各図に示している。

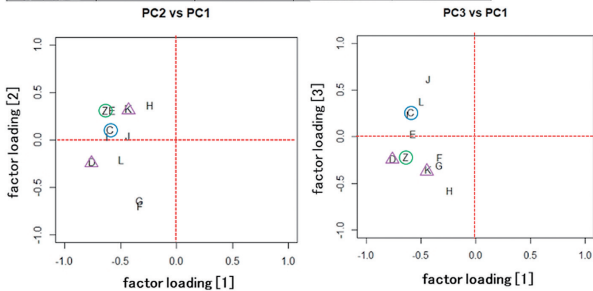
各図の表にはこれをもとに上記の手法で得た最適と見なしうる重回帰分析結果を示す。各因子負荷量プロット上では重回帰分析に用いた変数を色付けした○または△で囲んでいる（○は成果指標、△は評価点群または非評価点群の変数であることを示す）。

結果は以下のとおりである；

- ・最大CNCI（Z）は、国際共著論文割合（K）、CNCI論文数（C）、個別評価点（D）の3つの変数で全体の約42%の説明が可能であり、その寄与の大きさは概ね8：6：3である。
- ・合計CNCI（A）は、CNCI論文数（C）、国際共著論文割合（K）、重点的利用枠との共著論文割合（L）、研究計画の妥当性（H）の4つの変数で全体の約72%の説明が可能であり、その寄与の大きさは概ね4：2：1：1である。
- ・平均CNCI（B）は、国際共著論文割合（K）、研究体制の妥当性（I）の2つの変数を用いて全体の約25%の説明に留まる。その寄与の大きさは概ね2：1である。
- ・CNCI論文数（C）は、課題参加者数（J）、国際共著論文割合（K）、重点的利用枠との共著論文割合（L）の3つの変数で全体の

(a)

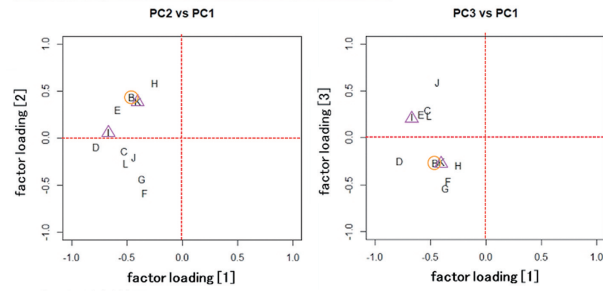
hp12~hp18A	目的変数	説明変数		
	最大CNCI (Z)	国際共著論文割合 (K)	CNCI論文数 (C)	個別評価点 (D)
重相関 R	0.6471			
重決定 R ²	0.4188			
補正 R ²	0.4033			
標準誤差	1.6415			
有意F	0.0000			
偏回帰係数		2.0438	0.3243	7.1114
標準化係数		0.4118	0.3232	0.1684
P値		0.0000	0.0000	0.0265
標準偏差	2.1160	0.4264	2.1086	0.0501



- 重回帰分析結果
- ◆ 約42%の説明が出来る。
- ◆ 説明変数は国際共著論文割合(K)、CNCI論文数(C)、個別評価点(D)
- ◆ それらの寄与の大きさは概ね 8:6:3

(c)

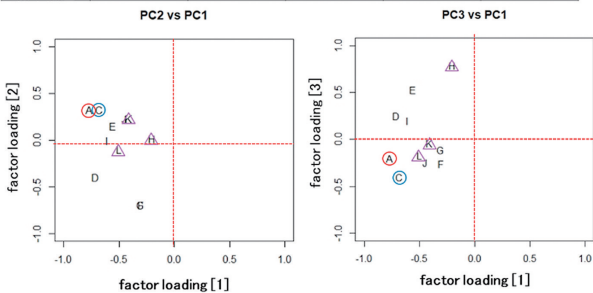
hp12~hp18A	目的変数	説明変数	
	平均CNCI (B)	国際共著論文割合 (K)	研究体制の妥当性 (I)
重相関 R	0.4987		
重決定 R ²	0.2487		
補正 R ²	0.2356		
標準誤差	1.1032		
有意F	0.0000		
偏回帰係数		1.2776	0.8172
標準化係数		0.4335	0.2083
P値		0.0000	0.0120
標準偏差	1.2564	0.4264	0.3202



- 重回帰分析結果
- ◆ 約25%の説明に留まる。
- ◆ 説明変数は国際共著論文割合(K)、研究体制の妥当性 (I)
- ◆ それらの寄与の大きさは概ね 2:1

(b)

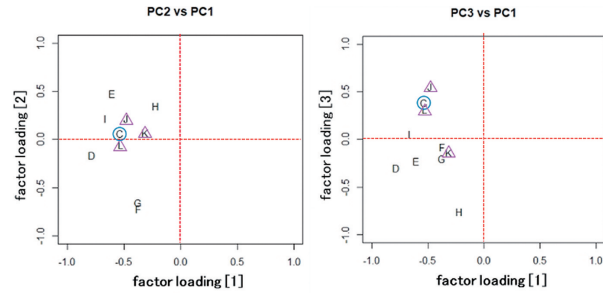
hp12~hp18A	目的変数	説明変数			
	合計CNCI (A)	CNCI論文数 (C)	国際共著論文割合 (K)	重点的利用枠との共著論文割合 (L)	研究計画の妥当性 (H)
重相関 R	0.8504				
重決定 R ²	0.7232				
補正 R ²	0.7134				
標準誤差	2.4897				
有意F	0.0000				
偏回帰係数		1.4934	3.2057	1.7243	1.6935
標準化係数		0.6801	0.2952	0.1464	0.1272
P値		0.0000	0.0000	0.0052	0.0130
標準偏差	4.6304	2.1086	0.4264	0.3932	0.3478



- 重回帰分析結果
- ◆ 約72%の説明が出来る。
- ◆ 説明変数はCNCI論文数(C)、国際共著論文割合(K)、重点的利用枠との共著論文割合(L)、研究計画の妥当性(H)
- ◆ それらの寄与の大きさは概ね 4:2:1:1

(d)

hp12~hp18A	目的変数	説明変数		
	CNCI論文数 (C)	課題参加者数 (J)	国際共著論文割合 (K)	重点的利用枠との共著論文割合 (L)
重相関 R	0.4127			
重決定 R ²	0.1703			
補正 R ²	0.1483			
標準誤差	1.9544			
有意F	0.0001			
偏回帰係数		0.0929	1.0105	1.0556
標準化係数		0.2468	0.2043	0.1969
P値		0.0059	0.0197	0.0262
標準偏差	2.1086	5.6024	0.4264	0.3932



- 重回帰分析結果
- ◆ 約17%の説明に留まる。
- ◆ 説明変数は課題参加者数 (J)、国際共著論文割合(K)、重点的利用枠との共著論文割合(L)
- ◆ それらの寄与の大きさは概ね 5:4:4

図11 「京」一般課題 (CNCI論文有) 重回帰分析結果。目的変数：(a) 最大CNCI、(b) 合計CNCI、(c) 平均CNCI、(d) CNCI論文数。図中の記号 (Z, A~L) の意味は図7の対応表を参照

約17%の説明に留まる。その寄与の大きさは概ね 5 : 4 : 4 である。

このように、成果の注目度の総量を反映する合計CNCI (A) で最も良好な回帰特性が得られた(約0.72(決定係数、以下同様))。最

大の寄与がCNCI論文数 (C) であることから、ある程度のCNCI値の論文が揃っていることを示している。次いで国際共著論文割合 (K) の寄与が大きい。寄与は小さいが重点的利用枠との共著論文割合 (L) や研究計画の妥当性 (H) も影響していることが明らかに

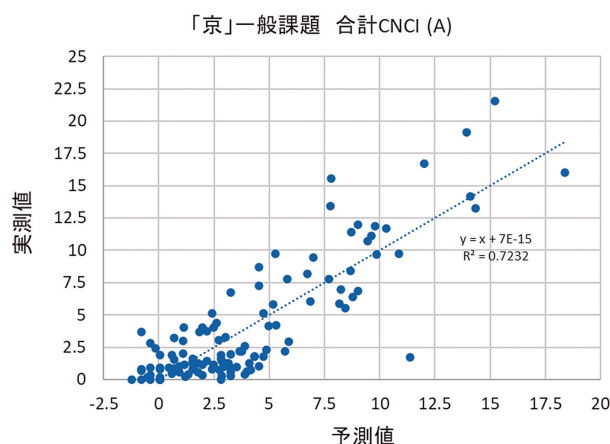


図12 合計CNCIの予測値と実測値の相関（「京」一般課題）

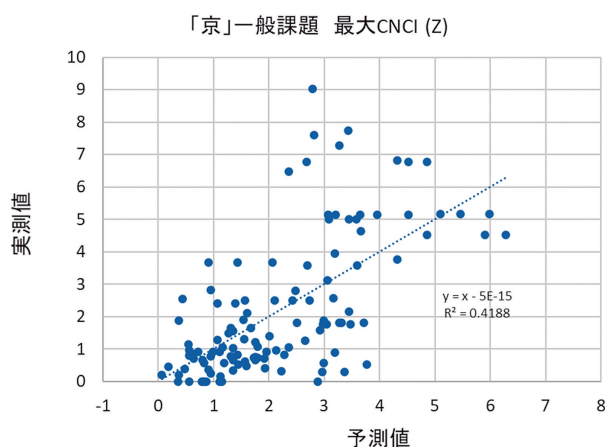


図13 最大CNCIの予測値と実測値の相関（「京」一般課題）

なった。合計CNCIの上記の説明変数を用いた予測値と実測値との相関を図12に示す。

一方、成果の注目度を反映する最大CNCI (Z) の回帰特性は合計CNCI (A) に比べるとあまりよくない(約0.42)。説明変数の組み合わせを合計CNCI (A) と比較すると、国際共著論文割合 (K) とCNCI論文数 (C) は同じであるが、順序が入れ替わっている(国際共著論文割合(K)の寄与の方が大)。とは言えCNCI論文数 (C) が多いほど最大CNCI (A) が向上する傾向があることが分かった。また、合計CNCI (A) の他の説明変数である研究計画の妥当性 (H) は最大CNCIでは個別評価点 (D) に置き換わり、その寄与の大きさも相対的に上がっている。合計CNCI (A) のもう一つの説明変数であ

る重点的利用枠との共著論文割合 (L) は最大CNCIではP値が大きいため (0.18) 外れている。最大CNCIの上記の説明変数を用いた予測値と実測値との相関を図13に示す。

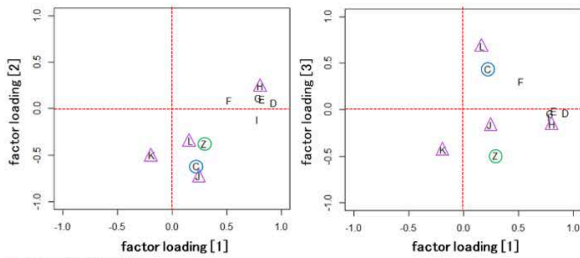
9.2. HPCI一般課題の場合

HPCI一般課題(CNCI論文有の課題に限定)の結果を図14 (a) ~ (d) に示す。目的変数は「京」一般課題と同様に、(a) が最大CNCI、(b) が合計CNCI、(c) が平均CNCI、(d) がCNCI論文数である。PCAの変数の組み合わせは「京」一般課題の場合と同様である。

- 結果は以下のとおりである；
- ・最大CNCI(Z)は、CNCI論文数 (C)、国際共著論文割合 (K)、課題参加者数 (J)、研究計画の妥当性 (H)、重点的利用枠との共著

(a)

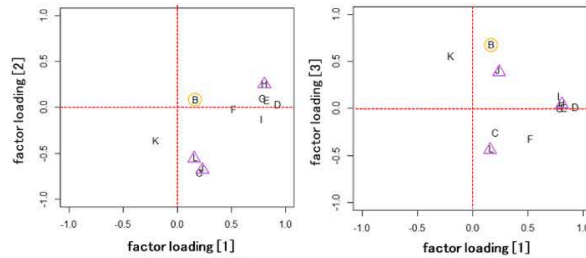
hp12~hp18	目的変数	説明変数				
	最大CNCI (Z)	CNCI論文数 (C)	国際共著論文割合 (K)	課題参加者数 (J)	研究計画の妥当性 (H)	重点的利用枠との共著論文割合 (L)
重相関 R	0.3507					
重決定 R2	0.1230					
補正 R2	0.1019					
標準誤差	1.6385					
有意F	0.0000					
偏回帰係数		0.0983	0.5958	0.0768	0.7888	-0.6336
標準化係数		0.1569	0.1405	0.1304	0.1273	-0.0712
P値		0.0268	0.0398	0.0633	0.0007	0.0586
標準偏差	1.7249	2.7524	0.4069	2.9267	0.5009	0.3489



- 重回帰分析結果
 - ◆ 約12%の説明に留まる。
 - ◆ 説明変数はCNCI論文数(C)、国際共著論文割合 (K)、課題参加者数 (J)、研究計画の妥当性 (H)、重点的利用枠との共著論文割合 (L)
 - ◆ それらの寄与の大きさは概ね 2.2:2.2:-1

(c)

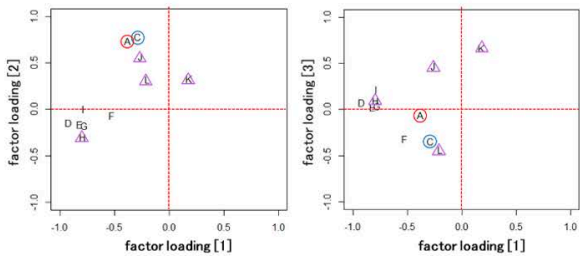
hp12~hp18	目的変数	説明変数		
	平均CNCI (B)	研究計画の妥当性 (H)	課題参加者数 (J)	重点的利用枠との共著論文割合 (L)
重相関 R	0.2296			
重決定 R2	0.0527			
補正 R2	0.0392			
標準誤差	1.2340			
有意F	0.0097			
偏回帰係数		0.3601	0.0544	-0.4838
標準化係数		0.1436	0.1269	-0.1344
P値		0.0340	0.0621	0.0479
標準偏差	1.2560	0.5009	2.9267	0.3489



- 重回帰分析結果
 - ◆ 約5%の説明に留まる。
 - ◆ 説明変数は研究計画の妥当性 (H)、課題参加者数 (J)、重点的利用枠との共著論文割合 (L)
 - ◆ それらの寄与の大きさは概ね 1:1:-1

(b)

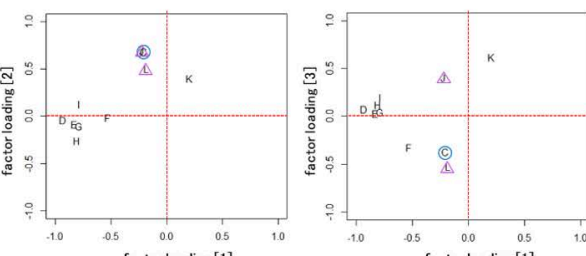
hp12~hp18	目的変数	説明変数				
	合計CNCI (A)	CNCI論文数 (C)	研究計画の妥当性 (H)	国際共著論文割合 (K)	課題参加者数 (J)	重点的利用枠との共著論文割合 (L)
重相関 R	0.7154					
重決定 R2	0.5118					
補正 R2	0.5001					
標準誤差	2.1999					
有意F	0.0000					
偏回帰係数		0.7514	1.1293	0.9327	0.1198	-0.7356
標準化係数		0.6663	0.1822	0.1223	0.1130	-0.0827
P値		0.0000	0.0003	0.0167	0.0312	0.1016
標準偏差	3.1042	2.7524	0.5009	0.4069	2.9267	0.3489



- 重回帰分析結果
 - ◆ 約51%の説明が出来る。
 - ◆ 説明変数はCNCI論文数(C)、研究計画の妥当性 (H)、国際共著論文割合(K)、課題参加者数(J)、重点的利用枠との共著論文割合(L)
 - ◆ それらの寄与の大きさは概ね 7.2:1.5:1.5:-1

(d)

hp12~hp18	目的変数	説明変数	
	CNCI論文数 (C)	課題参加者数 (J)	重点的利用枠との共著論文割合 (L)
重相関 R	0.3709		
重決定 R2	0.1375		
補正 R2	0.1294		
標準誤差	2.5742		
有意F	0.0000		
偏回帰係数		0.2453	1.8798
標準化係数		0.2608	0.2383
P値		0.0001	0.0003
標準偏差	2.7524	2.9267	0.3489



- 重回帰分析結果
 - ◆ 約14%の説明に留まる。
 - ◆ 説明変数は課題参加者数(J)、重点的利用枠との共著論文割合(L)
 - ◆ それらの寄与の大きさは概ね 1:1

図14 HPCI一般課題 (CNCI論文有) 重回帰分析結果。目的変数：(a) 最大CNCI、(b) 合計CNCI、(c) 平均CNCI、(d) CNCI論文数。図中の記号 (Z, A~L) の意味は図8の対応表を参照

論文割合 (L) の5つの変数が関与するが全体の約12%の説明に留まる。その寄与の大きさは概ね 2 : 2 : 2 : 2 : -1 である。

合計CNCI (A) は、CNCI論文数 (C)、研究計画の妥当性 (H)、国際共著論文割合 (K)、課題参加者数 (J)、重点的利用枠との共著論文割合 (L) の5つの変数で全体

の約51%の説明が可能であり、その寄与の大きさは概ね 7 : 2 : 1.5 : 1.5 : -1 である。

平均CNCI (B) は、研究計画の妥当性 (H)、課題参加者数 (J)、重点的利用枠との共著論文割合 (L) の3つの変数を用いて全体の約5%の説明に留まる。その寄与の大きさは概ね 1 : 1 : -1 である。

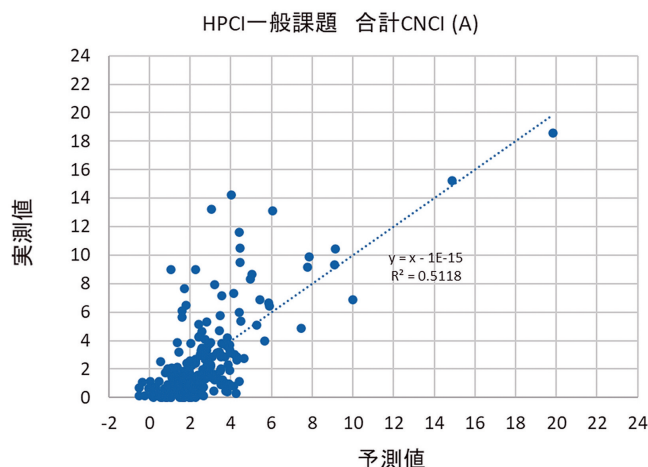


図15 合計CNCIの予測値と実測値の相関（HPCI一般課題）

- ・ CNCI論文数（C）は、課題参加者数（J）、重点的利用枠との共著論文割合（L）の2つの変数で全体の約14%の説明に留まる。その寄与の大きさは概ね1：1である。

成果の注目度の総量を反映する合計CNCI（A）の回帰特性のレベルは「京」一般課題の場合より低下する（約0.51）。説明変数の組み合わせは「京」一般課題の場合の4つ（CNCI論文数（C）、研究計画の妥当性（H）、国際共著論文割合（K）、重点的利用枠との共著論文割合（L））に課題参加者数（J）が加わっている。また、重点的利用枠との共著論文割合（L）はマイナスの寄与となっている。合計CNCIの上記の説明変数を用いた予測値と実測値との相関を図15に示す。

成果の注目度を反映する最大CNCI（Z）の回帰特性は「京」一般課題の場合に比べてかなり低下する（約0.12）。説明変数の組み合わせは「京」一般課題の場合の国際共著論文割合（K）、CNCI論文数（C）は同じであるが、「京」一般課題の場合の個別評価点（D）が研究計画の妥当性（H）に変わり、さらに課題参加者数（J）、重点的利用枠との共著論文割合（L）が加わる（後者はマイナスの寄与）。HPCI一般課題の合計CNCI（A）と比較すると、説明変数の組み合わせは同じであるが、寄与

の大きさのレベルが大きく変わり、CNCI論文数（C）、国際共著論文割合（K）、課題参加者数（J）、研究計画の妥当性（H）がほぼ同じレベルに並んでいる。

10. まとめ

10.1. HPCI成果発表データベースの進展

2013年度のシステム構築以降、受賞実績の登録と表示、研究分野別表示モードの導入、DOIデータベースからのインポート機能、CSV一括登録機能の追加など、利用者視点に立った改良を行ってきた。

2019年度までの全発表カテゴリの登録総数は12,442件（うちHPCI計算機資源は8,430件）、うち査読付き論文は3,250件（うちHPCI計算機資源は2,028件）に達する（2020/6/11現在）。

査読付き論文（英文）についてWOSとの照合を行い、信頼性の向上を図るとともにWOSにあって成果発表DBに登録漏れの論文を抽出し登録依頼を行うことにより登録件数の充実を行っている。

10.2. 成果論文のデータ分析

2012年度～2018年度実施の「京」一般課題、HPCI一般課題の全体（半年課題を除く）を対象に、成果論文のCNCI（注目度）に基づ

く成果指標（4種：最大CNCI、合計CNCI、平均CNCI、CNCI論文数）と、選定時の評価点群（6種）及び評価に直接関係しない国際共著論文割合等の非評価点群（3種）からなる説明変数（全9種）との間で相関分析、主成分分析、回帰分析を行った。

10.2.1 課題の種類単位の成果指標の統計値

「京」一般課題とHPCI一般課題につき、成果指標の内、CNCI値に係る3指標（最大CNCI、合計CNCI、平均CNCI）の各CNCI論文有課題全体での平均値を示した。成果の注目度を反映する最大CNCIの平均値は「京」一般課題が約2.8、HPCI一般課題は約1.7、成果の注目度の総量を反映する合計CNCIの平均値は「京」一般課題が約4.8、HPCI一般課題は約2.9、成果の平均的な注目度を反映するCNCIの全体平均値は「京」一般課題が約1.7、HPCI一般課題は約1.1である。CNCI = 1は世界平均と同等のパフォーマンスを表すが、「京」一般課題、HPCI一般課題とも、最大CNCIの平均値はこれを十分上回り、CNCIの全体平均値も1を上回っている。

3成果指標の度数分布はいずれもやや平坦なべき乗分布をしている。特に、合計CNCIの場合は「京」一般課題、HPCI一般課題ともすそ野が高い値にまで伸びている。各成果指標が平均値の課題は「京」一般課題、HPCI一般課題とも全体の上位3割程度に位置する。

一方、CNCI論文数の平均値は「京」一般課題とHPCI一般課題で同程度であり、約2.8である。

10.2.2 成果指標と説明変数の間の相関

「京」一般課題の場合は成果指標と国際共著論文割合（非評価点群）、研究計画の妥当性（評価点群）との間の相関が比較的強い。また、成果指標と個別評価点（評価点群）、科学的・社会的意義（同）、研究体制の妥当性（同）との間の相関、及び成果指標と重点的利用枠

との共著論文割合（非評価点群）との間の相関も認められる。

HPCI一般課題の場合は成果指標と課題参加者数（非評価点群）との間に比較的強い相関がある。また、CNCI論文数を除く成果指標と国際共著論文割合（非評価点群）との間の相関、CNCI論文数を除く成果指標と全評価点群との間の相関、及びCNCI論文数と重点的利用枠との共著論文割合（非評価点群）との間の相関も認められる。

10.2.3 成果指標の回帰分析

「京」一般課題の場合は、成果の注目度の総量を反映する合計CNCIで最も良好な回帰特性（約0.72（決定係数、以下同様））が得られた。最大の寄与がCNCI論文数、次いで国際共著論文割合である。寄与は小さいが重点的利用枠との共著論文割合や研究計画の妥当性も影響している。成果の注目度を反映する最大CNCIの回帰特性は合計CNCIに比べると余り良くない（約0.42）。説明変数の組み合わせは国際共著論文割合、CNCI論文数、個別評価点である。寄与の大きさは国際共著論文割合がCNCI論文数をやや上回る。個別評価点の寄与の大きさは合計CNCIに於ける研究計画の妥当性より相対的にやや高い。

HPCI一般課題の場合は、合計CNCIの回帰特性のレベルは「京」一般課題の場合より低下する（約0.51）。説明変数の組み合わせは「京」一般課題の合計CNCIの4つに課題参加者数が加わる。最大の寄与がCNCI論文数、次いで研究計画の妥当性、次いで国際共著論文割合、課題参加数である。重点的利用枠との共著論文割合はマイナスの寄与となる。最大CNCIの回帰特性は「京」一般課題の場合に比べてかなり低下する（約0.12）。説明変数の組み合わせは合計CNCIと同じであるが、寄与の大きさはCNCI論文数、国際共著論文割合、課題参加者数、研究計画の妥当性でほぼ横並びとなる。

11. おわりに

2012年度より開始された「京」を中核とするHPCI計算機資源利用研究に係る成果論文のデータ分析を進めて来た。課題の成果論文の注目度を反映する最大CNCIの平均値の評価により、CNCI論文有の課題全体として世界平均を上回るインパクトのある研究成果を上げたことを示すことが出来た。課題の成果の注目度の総量を反映する合計CNCIの回帰分析から、これに寄与する説明変数にはCNCI論文数、国際共著論文割合に加え、課題選定時の評価項目（研究計画の妥当性）も含まれることが明らかになり、課題選定の妥当性を支持する分析結果も得られた。

今後引き続きデータの蓄積、分析を継続していく必要がある。得られた知見により課題選定方式の更なる改善、ひいてはすぐれた成果創出の更なる拡大につながることを期待するものである。

謝辞

本報をまとめるに当たり、全編にわたり有益な助言をいただいた鈴鹿医療科学大学学長豊田長康氏に深く感謝致します。HPCI成果発表DBの運用に当たりデータの提供にご協力いただいた全てのHPCI利用研究課題の課題代表者／連絡責任者／課題参加者／関係者の方々に感謝致します。

参考文献

- [1] HPCIポータルサイト, [HPCIの概要], [https://www.hpci-office.jp/pages/what_is_hpci].
- [2] 文部科学省ホームページ, [告示・通達], [https://www.mext.go.jp/b_menu/hakusho/nc/_icsFiles/afiedfile/2019/09/24/1421213_005_1.pdf].
- [3] HPCIポータルサイト, [成果発表データベース], [https://www.hpci-office.jp/hpcidatabase/publications/search.html].
- [4] 木村晴行、平塚篤; 「京」を中核とするHPCI成果発表データベースの構築」 RIST NEWS, No.56 (2014年1月) 2-13.
- [5] XSEDE USER PORTAL, Publications, [https://portal.xsede.org/publications/#].
- [6] INCITE, [https://www.doeleadershipcomputing.org/].
- [7] PRACE, [https://prace-ri.eu/?lang=en].
- [8] InCites Benchmarking, [https://clarivate.jp/products/incites/].
- [9] CNCI, [http://help.incites.clarivate.com/inCites2Live/indicatorsGroup/aboutHandbook/usingCitationIndicatorsWisely/normalizedCitationImpact.html].
- [10] 文部科学省 科学技術・学術政策研究所, 「科学研究のベンチマーキング2019 -論文分析でみる世界の研究活動の変化と日本の状況-」, 2019年8月.
- [11] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [https://www.R-project.org/].
- [12] 尾崎幸洋・赤井俊雄・宇田明史『化学者のための多変量解析—ケモメトリックス入門』講談社サイエンティフィック, 2002年.