Monitoring Progress in the Transition to Open Access

Report of a Working Group

March 2014

Contents

Executive Summary

1. Introduction
2. General principles and coverage
3. Accessibility
4. Accessibility: the availability of OA options
5. Usage
6. Sustainability
7. Quality of services
8. Costs and governance

Annexes

A. Members of the working group
B. Criteria for identifying versions
C. Moving cohort samples for articles
D. Building a stratified sample of universities
E. Costs

Executive Summary

The Finch Report[1] urged that the process to implement its recommendations should include the collection of reliable, high-quality indicators on the key features of the changing research communications landscape. The precise configuration of the indicators, data definitions, protocols and methodologies should, the Report suggested, be considered by representatives of the different stakeholder groups. This current report presents proposals for a framework of indicators to monitor progress towards open access (OA) in the UK.

The proposals do not attempt to cover all elements of the Finch recommendations, but focus on the transition to OA. Our aim has been to produce a framework of indicators which address key questions relating to the transition, based on data that can be gathered annually in relatively straightforward fashion. The picture the indicators present should be reasonably authoritative. But we recognise that some important questions require more detailed research; and we have concluded reluctantly that these must for the present lie outside the scope of the framework we propose. All our proposals, however, including our exclusions from the framework, should be kept under review in the light of experience and of developments in the availability of data.

We have built our proposals around the core themes of the Finch Report: accessibility, sustainability and quality of service. We also urge that the work to produce the indicators should be undertaken annually in order to monitor and assess trends. Clarity and consistency in definitions and approaches to data collection and analysis are therefore critically important. It is also important that, with as few exceptions as possible, all the results of the data gathering and analysis should be made freely available for others to interrogate and to use.

The work we propose should be overseen by representatives of the key stakeholder groups, including funders, universities and other research organisations, libraries, Jisc, learned societies, and publishers. UniversitiesUK is now taking the lead in convening a body comprising such representatives, and we hope it will take on that role. Once the new body is established, we envisage that it will wish to consider our proposals, and the priorities to be attached to our various proposals and recommendations, in detail.

Proposals

Accessibility

1. There should be an annual exercise to assess the numbers – and the proportions of the overall totals – of all articles and of those with a UK author that are accessible free of charge from
    a. fully OA journals
    b. hybrid journals
    c. journals that provide free access on their platforms after an embargo period
    d. repositories and other websites

[1] Working Group on Expanding Access to Published Research Findings *Accessibility, Sustainability, Excellence: how to expand access to research publications*. June 2012. Available at http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf

2. The counts of articles should distinguish between pre-prints, authors' accepted manuscripts, and published versions of record
3. The counts should include all articles found to be accessible free of charge, whether they have been posted illicitly or not; but an estimate – based on checking a sub-sample of them – should be produced of the numbers of illicitly-posted articles
4. The counts should be based on automated searches for samples of the articles recorded in either the SCOPUS or CrossRef databases, plus a full census of those published in fully-OA journals
5. Searches should be made for at least four global and UK-authored samples of articles: those published 1,7, 13 and 25 months earlier.
6. For pragmatic reasons, the date of publication should be taken as the date of the relevant issue of the journal; but we **recommend** that all publishers should include the date of publication in the metadata for all articles.
7. The results of the counts should be broken down in accordance with the four subject panels established for the REF; but there should be no other breakdown

Availability of OA options

8. In order to monitor trends in the numbers of journals that comply with RCUK and Funding Bodies' policies, data should be gathered from the major publishers on the numbers of fully-OA and hybrid journals, the length of embargo periods, levels of APCs, and the availability of CC BY and other licences
9. There should be an audit check of the data supplied by publishers

Usage

10. It will be important at some point in the future to assess the extent to which OA results in higher levels of usage, especially by people and organisations beyond the university and research communities. But for the present, such studies should remain outside the initial framework of indicators we propose.

Financial sustainability

11. Aggregate data should be gathered annually from a stratified sample of universities on their expenditure on
    a. APCs and other publication charges, and the numbers of articles for which they have been paid
    b. journal subscriptions, including both print and e-journals
12. Ideally, information should also be gathered on universities' expenditure from Wellcome Trust and RCUK block grants to meet the costs of APCs and for other purposes; their use of other funds to meet APCs; and on the overall numbers of articles published, and the number deposited in the institutional repository.
13. Data on the numbers of titles (fully-OA, hybrid, delayed access and subscription-only) and of articles published in each category of journal should be gathered from the major publishers. An audit check of numbers should be undertaken on a sample of the journals

14. The survey we propose in Section 7 should include conjoint analysis to assess the extent to which various factors, including the price of APCs, influence authors' decisions on where to publish
15. Data should be gathered from a sample of UK learned societies on their overall as well as their journal-related income and expenditure.

Quality of services to authors and readers

16. Annual surveys should be undertaken to assess authors' perceptions and experiences in relation to their most recent completed interaction with a journal
17. We do not recommend at this stage a parallel survey of readers and users.

1. Introduction

1.1. The report of the Finch Group in 2013 noted that implementation of its proposals would be an intricate process that would require co-ordination and the active engagement of funders, universities, publishers and learned societies, as well as Government. It stressed the importance of the different groups continuing to work together; and it urged that in order to sustain the confidence of all parties and stakeholders, the implementation process should include the collection of reliable, high-quality indicators on the key features of the changing research communications landscape. The precise configuration of the indicators, data definitions, protocols and methodologies should, the report suggested, be considered by representatives of the different stakeholder groups.

1.2. In June 2013 the Research Information Network (RIN) convened, with the support of Jisc, the Wellcome Trust, the Publishers Association and Research Libraries UK, a working group of representatives of publishers, funders, learned societies and libraries to consider the configuration of a set of indicators that can be generated on a regular basis[2]. The aim has been to meet the requirement set by the Finch Group for reliable indicators that are generated in ways that secure the confidence of all parties. Our proposals are therefore founded on the need to monitor and assess trends in the transition to open access (OA) and to answer key questions from the varying perspectives of the different stakeholders.

1.3. We do not attempt to cover in our proposals all elements of the Finch recommendations. Our focus is on the transition to OA, and our approach to definitions of OA has been essentially pragmatic, as set out in Section 3 of this report. But we do not address other elements in the Finch recommendations relating to extensions to licensed access for the benefit of SMEs and other organisations, or the Access to Research pilot scheme recently launched to provide walk-in access to scholarly journals in public libraries[3]. We recognise, of course, the importance of monitoring progress on the scale and impact of these elements of the Finch recommendations; but we concluded that such monitoring could not be covered in the initial range of indicators that we outline in this report. As with other elements in our proposals, however, we recommend that this should be reviewed in the future[4].

1.4. Three further points should be made by way of introduction. First, we have concluded, not surprisingly, that some important questions (for instance, about the relationships, if any, between embargo periods, journal half-lives, and risks to journal subscriptions) cannot be answered by any straightforward creation of indicators from the available data. They require more substantial research projects which lie beyond the scope of the framework of indicators we propose, although we believe that some of the data collection that we recommend may serve as useful input into such projects. We point to a number of such questions in the body of the report.

---

[2] A list of members of the group is at Annex A

[3] See http://www.acesstoresearch.org.uk

[4] The current Access to Research pilot, led by the Publishers Licensing Society and the Society of Chief Librarians, is due to last for two years, ending in early 2016. We presume that some evaluation exercise will be undertaken before then, and that if the scheme continues, monitoring arrangements will need to be put in place.

1.5. Second, our proposals are based on our best assessment of the availability of data at present. In some cases (for instance in relation to the inclusion of information about licences and date of publication in standard metadata) we believe that the position will improve in the relatively near future, in response to work that is now under way on the part of universities, publishers and others, in part following recommendations from the National Information Standards Organization (NISO, an organisation based in the US) and other bodies. We have also sought to make our proposals as clear and specific as possible, recognising, for example, that data definitions and standards are of critical importance. We have no doubt, however, that our proposals will need to be amended as circumstances change and in the light of experience; and we recommend a light-touch review after the first annual exercise, and a more thoroughgoing review after the second. It is also important, however, that the reviews should be overseen by representatives of the different stakeholder groups, who should also agree any amendments to be made. This is an issue of governance which we also address in Section 6 of the report.

1.6. Third, we have built our proposals for indicators around the three core themes highlighted in the Finch Report: accessibility, sustainability, and excellence (which for the purposes of our work we have interpreted as quality of service). Much of our attention has focused on the first of these, accessibility. We do not under-estimate the importance of sustainability and quality of service; they are of critical importance to all stakeholders. Hence we make some proposals in those two areas. But we concluded that some of the most important questions in those areas need more detailed research than can be covered through the kind of regular data gathering exercises that form the substance of our proposals.

2. Aims, general principles and coverage

Aims

2.1. Our aim has been to identify indicators that can provide a reasonably authoritative picture of progress in the transition to OA in the UK: one that can be undertaken at reasonable cost and up-dated annually, based on consistent definitions and approaches to data collection and analysis. A subsidiary aim is that the results of the data gathering and analysis should be available for others to use, as the basis for further analysis and for any more detailed projects that they wish to undertake.

Principles

2.2. In order to meet these aims, five overlapping general principles underlie our proposals:

a. key questions: we have focused our attention on practical ways in which to gather and analyse data to address questions of importance to the range of stakeholders, from their varying but overlapping perspectives

b. joint ownership and responsibility: while individual organisations may take responsibility for providing the data and perhaps also for initial analysis, the work should be overseen by representatives of the key stakeholder groups, including funders, universities and other research organisations, libraries, Jisc, learned societies, and publishers. Universities UK is now taking the lead in convening a body comprising such representatives, and we hope it will take on that role, as well as publishing the results of the work.

c. openness: the definitions and protocols underlying the data and the analysis, together with the data and the results of the work, should be deposited and made accessible via a repository under a CC BY licence. We recognise, however, that in a few cases constraints may arise on grounds of confidentiality and competition law. These should be kept to a minimum.

d. cost-effectiveness and practicality: the framework of indicators does not seek to answer all possible questions, but to produce indicators on key issues where data can be readily gathered and analysed in a consistent fashion. This should become more straightforward as more data becomes routinely available.

e. regularity: in order to monitor and assess trends, it is crucial that the overall exercise we propose, in line with the Finch Group's report, should be repeated annually, and at the same point in the calendar year, in order to achieve as high a level of consistency as possible.

Coverage

2.3. In line with the OA policies of the Research Councils and Funding Bodies focus on articles published in scholarly journals, our proposals are restricted to data and indicators covering such publications. To be explicit, our proposals cover research articles, review articles and conference proceedings published in peer-reviewed journals with an ISSN number. They do not extend to other

kinds of publications such as books, book chapters, or data. Henceforth we shall use the term 'articles' to refer to all the publications within the scope of our proposals.

3. Accessibility

3.1. In order to assess the impact of new policies from funders and universities in the UK, it is clearly important to monitor trends in the numbers - and the proportions of the overall population - of articles published and accessible on OA terms. Moreover, those trends must be monitored both for the global population of articles, and for those which include in the list of authors at least one who has an affiliation to a UK institution. But our aim has been – with cost very much in mind - to set a framework for monitoring broad trends rather than specifications for detailed studies. Hence with the exception of the broad subject breakdown noted at paragraphs 3.15 and 3.19 below, we do not propose that data should be gathered on trends at a more detailed level relating to specific institutions or categories of institution, publishers or categories of publisher and so on. The indicators should be restricted to global and UK levels.

Routes to OA

3.2. We recognise that there are different routes to through which the increases in access recommended by the Finch Group can be achieved. We propose an annual exercise to assess – making use of automated search techniques that mirror those used by reasonably-competent researchers - the numbers and proportions of articles under four headings:

a. articles published in fully-OA journals, including not only those listed in the Directory of Open Access Journals (DOAJ), but also those published in any other journal that makes clear that it publishes on OA terms (see paragraph 4.2 which outlines an issue relating to delays in inclusion in the DOAJ). The count of articles should distinguish between those published in *journals* that do or do not charge an APC (there should be no attempt to assess whether or not an APC was in fact paid in individual cases).

b. articles published on OA terms in hybrid journals, and to which payment of an APC applies. Previous exercises[5] have noted significant numbers of articles that have been made accessible for promotional or related purposes, usually for a limited period, and are subsequently found not to be accessible free of charge. We propose in paragraph 3.18 a means of distinguishing between such articles and those for which an APC has been paid, which are thus accessible on OA terms in perpetuity .

c. articles published in subscription-based journals that nevertheless provide access free of charge on the publisher's platform after embargo periods that may range from 6 to 24 months. Since the maximum embargo period allowed under RCUK's and the Funding Councils' policies is 24 months, articles made accessible after an embargo of longer than 24 months should be excluded from the count. We are aware that some journals which operate in this way also provide a hybrid option so that articles are made accessible free of charge immediately on publication. Our research indicates that signals are provided via

[5] See, for example, Eric Archambault et al, *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011*, Montreal, Science Metrix, 2013.

tables of contents and other means it is usually possible readily to identify such articles, even after the embargo period has expired.

d. articles published in subscription-based journals, but where a version of the article - which may range from the version that was submitted to the journal, through an authors' accepted manuscript (AAM) (thus including changes made as a result of peer review), up to and including the final published version – is made accessible via authors' web pages, an institutional or subject-based repository, or some other website. Funders have set specific requirements relating to the versions of articles they require to be made accessible, and on which sites. We address issues relating to versions in paragraph 3.6.

3.3. The count for category a) above will be based on a census of all the articles in the listed journals. The count for categories b) c) and d) will be based on automated searching for samples of articles (see paragraphs 3.12-3.18 below). The automated search will also find articles in category a), but these should be de-duplicated in overall counts which should be based on the census.

3.4. The overall count should distinguish between versions of articles found under the four heads, and any sum total of all articles found to be accessible free of charge should be subject to careful de-duplication, so that no individual article is counted more than once in any total sum, simply because it is found in multiple places and versions.

3.5. With regard to articles found to be accessible (as in 3.2.d above) from repositories or other sites - any site other than the publishers' – two issues require careful handling.

3.6. First, the count should seek to distinguish, using techniques we outline in Annex B, between different versions[6]:

a. those which show no sign of having been subject to peer review, which may be taken to be a version up to and including that which was submitted to a journal (often termed a 'preprint')

b. those which show clear indications of having been subject to peer review, but which are clearly not the final version as published by the journal. Although in a few cases this may represent an intermediate stage before final acceptance, in the great majority of cases this is likely to represent what is often termed an 'accepted author manuscript' (which is the version to which access is required to meet the terms of RCUK's policies and those of many other funding bodies).

---

[6] We are aware that the categorisation below is simpler, and thus cruder, than the six categories of version defined by ALPSP and NISO.(See *Recommendations of the NISO/ALPSP Working Group on Versions of Journal Articles* , 2006, available at http://www.niso.org/apps/group_public/download.php/48/Recommendations_TechnicalWG.pdf). We also recognise that versions that that have been subject to peer review (often referred to as Authors' Accepted Manuscripts) may vary significantly in terms of the extent of variances from the final published version of record. Nevertheless, we believe the typology we set out here is adequate for the purposes of the monitoring framework we propose.

c. the published version of record (usually termed either 'published journal article' or 'version of record').

3.7. Second, we are aware that different versions of some articles are made freely accessible – on authors' web pages but also via other sites and services - in ways that may contravene the terms and conditions set by individual journals and publishers, particularly with regard to embargo periods and the versions that are made accessible. Many such articles would be found by researchers and others using such services as Google Scholar, and it would be wrong to exclude them from the count. But the key focus in monitoring, we suggest, should be on trends in licit rather than illicit behaviours[7]. In order to check whether changes in the scale of illicit behaviours (or attempts to stop them) are having an effect on the overall totals, we propose that a sub-sample of the versions of articles found on sites other than the original publisher's (thus those in category d) set out in paragraph 3.2 above) should be checked against the current policies of the relevant journals. This will produce a rough estimate of the overall extent of illicit postings, and, over time, of any changes in the impact they may be having on the overall count of articles that are freely accessible. Since the estimate will cover only a sample of articles, no attempt should be made at any breakdown by publisher, institution or author.

Data sources

3.8. It is estimated that over 2 million peer-reviewed articles are published annually, and that over 130k of those have at least one UK author. A report recently commissioned by BIS[8] was based on the SCOPUS database. It found that that nearly 8% of UK articles in 2012 were published in wholly OA journals, and that author accepted manuscripts were accessible via repositories or other websites for over 11% of UK articles.

3.9. We have considered various sources of data – including Ulrichsweb, Web of Knowledge and KB+ - for determining the overall populations of journals and articles published globally, along with those articles published UK authors. We have concluded that CrossRef and SCOPUS provide the fullest coverage, and that one or other of them should be used for the purposes of the framework: determining the overall populations, and also for creating the samples of articles to be searched for. We are aware, of course, that journals and articles are published that appear in neither database; and that each changes in composition over time. The decision on which of the two to use should be made on pragmatic grounds. But in order to achieve consistency in assessing trends over time, a change from one year to another should be avoided if possible.

Timing of publication

3.10. Our research has shown a disturbing lack of evidence and consistency in practice in recording the date of publication for articles. The ways in which the date of publication is recorded varies hugely across journals, and even as between what is shown on different versions of Tables of Contents, and

---

[7] We recognise that what counts as licit or illicit behaviours may change as the policies adopted by journals change.
[8] *International Comparative Performance of the UK Research Base - 2013*. A Report prepared by Elsevier for the UK's Department of Business Innovation and Skills, 2013

between HTML and PDF versions of articles. Some journals show on some versions of articles information such as dates of submission, acceptance and online publication. But the latter date does not always relate to the final published version; and the date when an article was finally published in an issue of the journal may not be shown at all. Yet other journals do not provide any information of this kind, and the various versions of articles may show no more than a year of publication. The starting date for embargo periods is thus in some cases wholly unclear. We therefore **recommend** that all journal publishers should take steps to ensure that the date of publication – where relevant, the one that they use as the starting date for embargo periods - is always included in the standard metadata for articles[9].

3.11. The SCOPUS database includes information on the publication date for each issue of the vast majority of the journals it covers; and that is probably the most reliable and consistent date to use for the present. We recognise that even issue dates are problematic in some cases, particularly for journals that publish relatively infrequently (say, on a quarterly or half-yearly basis). It is not uncommon for journals to publish, say, the expected March issue in February or in April, but for the stated date of publication still to be March. The incidence of such quirks is unlikely to change over time, however, and since our focus is on trends, we **propose** that the date of issue should be used as the basis for sampling and searching for the present, unless and until journals adopt more consistent ways of recording the date of publication.

Sampling

3.12. The count we propose of articles in fully-OA journals should be based on a full census of such journals. But we have concluded that in order to achieve our objective of monitoring trends in the numbers and proportions of articles accessible under the other three heads outlined in paragraph 3.2 above, automated searches should be undertaken annually for at least four samples of the overall population of articles: those published 1,7,13 and 25 months previously[10]. These periods correspond to the policies of RCUK and the Funding Councils, which prescribe access free of charge, if not immediately, then after maximum embargo periods of 6, 12 or 24 months. In each case, we propose searching after month's grace to allow for practical issues which may arise in depositing or providing access to articles.

3.13. In order to assess trends relating both to the global population of articles, and to those with a UK author, two sets of the four samples should be created: for the UK and for the world. The sampling strategy will also have to take some account of disciplinary differences.

3.14. The numbers of articles published each year varies hugely across disciplines: the SCOPUS database records, for instance, over 600k articles published globally in the clinical sciences, but only 80k in the humanities. Moreover, differences in the speed and scale of the transition to OA in different subject

---

[9] This is in line with the recommendations of the NISO working group on Open Access Metadata and Indicators, issued for consultation in January 2014

[10] If in time it is judged that automated validation of the findings of automated searches, and identification of the different versions of articles, is good enough to preclude the need for manual checking, then sampljng may no longer be required, and counts could be based on the whole population of articles published.

areas have frequently been noted; and they are reflected in the policies of RCUK and the UK Funding Bodies.

3.15. We propose, therefore, that samples should be constructed in four broad subject areas: life sciences and medicine, physical sciences and engineering, social sciences, and arts and humanities. These correspond broadly to the main panels constituted for the REF. The allocation of articles to subjects should be based on the subject category of the journal in which it was published, although a judgement will be required for articles published in one of the large multidisciplinary journals.

3.16. The size of the UK and global samples in the four subject areas should be large enough to provide for reasonable margins of error. But it should be small enough to allow for manual verification of the results of the searches, and to determine the versions of articles found to be accessible. Experience suggests that an average of 5-6 versions of articles will be found for each article in the samples. These will include nil returns, but also duplicates of all three versions outlined above, as well as versions of working papers, presentations etc that may be related to but are not the same as the article itself.

3.17. There will of course be a trade-off between sample size and cost in terms of the time to be spent in manual verification of the search results. Initial advice suggests that samples of c400 for each subject area should suffice[11].

3.18. We have considered whether the samples of articles to be searched for should be created anew for each annual exercise, or whether renewed searches should be undertaken for samples already created for previous exercises. New samples for each exercise have the attraction of simplicity and accuracy in tracking the current uptake of OA. But a cohort sampling approach may have potential for greater accuracy in tracking the length of the delays before articles become accessible, and the scale of 'promotional access' (see paragraph 3.2 b above). Annex C illustrates a schedule for creating and searching for moving cohort samples.

Subject and other breakdowns

3.19. As noted above, we do not propose any further subject breakdowns, or any analysis of differences between publications from authors affiliated to institutions of different types, or between articles published in different categories of journals.

Licence terms

3.20. RCUK's policy seeks to promote re-use of material presented in published articles, as well as access for reading free of charge. It therefore requires the use of CC-BY licences when an APC is paid, and licences with as few restrictions as possible when access is provided through other routes. It is important, therefore, that the use of different licences should be tracked as far as is reasonably possible.

3.21. There is some way to go before publishers adopt standard methods for recording licence terms in the metadata for articles[12]. Where Creative Commons licences – CC BY but also other CC licences -

[11] This will give a confidence level of 95% and an error margin of 4-5%.

are properly used (though that is not necessarily the case at present), it should eventually be straightforward to identify them automatically in different versions of articles. But not all publishers employ CC licences, and for the different versions of articles found in repositories and other sites, it may be difficult to determine precisely which licence has been used. Any count will therefore not be fully-comprehensive. Nevertheless, it should prove useful in determining trends.

3.22. The introduction by CrossRef of a requirement for metadata to include information about licences should, once it is (as is likely) widely adopted, bring greater comprehensiveness and accuracy in the counting of different licences.

[12] We are aware, of course, of the recommendations of the NISO working group on Open Access Metadata and Indicators, issued for consultation in January 2014; but it may be some time before they are fully adopted.

Summary

- There should be an annual exercise to assess the numbers – and the proportions of the overall totals – of all articles and of those with a UK author that are accessible free of charge from
    - fully OA journals
    - hybrid journals
    - journals that provide free access on their platforms after an embargo period
    - repositories and other websites
- The counts of articles should distinguish between pre-prints, authors' accepted manuscripts, and published versions of record
- The counts should include all articles found to be accessible free of charge, whether they have been posted illicitly or not; but an estimate – based on checking a sub-sample of them – should be produced of the numbers of illicitly-posted articles
- The counts should be based on automated searches for samples of the articles recorded in either the SCOPUS or CrossRef databases and, plus a full census of those published in fully-OA journals
- Searches should be made for at least four global and UK-authored samples of articles: those published 1,7, 13 and 25 months earlier.
- For pragmatic reasons, the date of publication should be taken as the date of the relevant issue of the journal; but we **recommend** that all publishers should include the date of publication in the metadata for all articles.

4. Accessibility: the availability of OA options

4.1. In addition to monitoring the numbers and proportions of articles that are accessible on OA terms, it is also important to monitor on a systematic basis trends in the numbers of fully-OA and hybrid journals, and in particular those that offer options to authors that are compliant with RCUK and Funding Bodies' policies.

4.2. The key source for the numbers of fully-OA journals should be the DOAJ. But we have noted that – even before the DOAJ tightened its procedures for determining whether journals should be included in its listings – there can be significant delays before new journals are included. We therefore propose that the DOAJ count should be supplemented by a note of evidence from major publishers and learned societies of fully-OA journals that are yet to be listed in DOAJ.

4.3. There has also been widespread speculation on three further issues relating to trends in the 'supply' of OA options, which it would be useful to monitor over time:

a. the length of embargo periods prescribed by publishers of different journals and whether they are tending to rise or fall

b. the level of APCs, and again whether they are tending to rise or fall

c. the extent to which publishers allow for CC BY licence terms

4.4. Data on the length of embargo periods and on licence terms is gathered by SHERPA, although there has been some criticism that since the data has been up to now gathered essentially at publisher level rather than the level of individual journals, it is not as comprehensive and accurate as some commentators would wish. Moreover, until now, it has not been possible to track changes in the SHERPA database. Steps are now in hand to deal with both issues.

4.5. In the meantime, however, we propose that data on the numbers of fully-OA and hybrid journals – as well as embargo periods, APCs, and use of CC-BY and other licences - should be gathered from all the major publishers. An audit check should then be undertaken for a sample of the most popular journals in which UK authors have published over the previous year in each of the four subject groups we outline in paragraph 3.15.

Summary

- In order to monitor trends in the numbers of journals that comply with RCUK and Funding Bodies' policies, data should be gathered from the major publishers on the numbers of fully-OA and hybrid journals, the length of embargo periods, levels of APCs, and the use of CC BY and other licences
- There should be an audit check of the data supplied by publishers

5. Usage

5.1. One of the key motivations underlying OA policies is the belief that they will increase not only accessibility but use of the published findings of research, not only in the university and research sectors, but far beyond. Reporting on usage has developed considerably in accuracy and comprehensiveness as a result of the successive COUNTER protocols and codes, as well as the moves to gather COUNTER-compliant data from repositories. The latest version of the COUNTER code helps further by requiring OA articles in hybrid journals to be counted separately.

5.2. It would of course be useful to assess whether and to what extent OA content is associated with increases usage by members of demographic groups beyond the university and research communities. At present, however, it is not possible to gather systematic data on the demographics of users either on publisher platforms or via repositories. Detailed studies would therefore be required to answer any questions on the demographics of usage. For these and other reasons we have therefore concluded that it is not possible for the present to bring indicators of usage into the framework we propose; but the matter should be kept under review as new sources of data become available.

Summary

- It will be important at some point in the future to assess the extent to which OA results in higher levels of usage , especially by people and organisations beyond the university and research communities. But for the present, such studies should remain outside the framework of indicators we propose

6. Financial sustainability

6.1. The financial implications of the transition to OA are matters of concern to all parties. From the perspectives of researchers, universities and funders, there are concerns about the overall costs of the transition, including the continuing costs of subscriptions, the newer costs of APCs (along with other publication charges), and the costs of administration. From the perspectives of publishers and learned societies, there are concerns about the potential impact on their revenues from all sources: the extent to which increases in APC revenues will match potential falls in subscription income (as well as other sources of income from rights, advertising and so on), and again the administrative costs involved. Financial risks, and how they can be managed and minimised, are much on the minds of all parties.

6.2. In these circumstances, it is important that individual universities, funders, publishers and learned societies should have access as soon as possible to reliable information that will enable them to locate their own position within a wider context.

Trends in the amounts paid in subscriptions and APCs

6.3. Much attention has focused on the likely increases in overall expenditure on a growing number of publications as universities and funders meet the costs of APCs for articles published in fully-OA as well as hybrid journals, on top of the costs of subscriptions for wholly-subscription-based and hybrid journals. The scale of any increase is unknown; but it will change over time and will vary considerably across different institutions. If we imagine ourselves into a wholly OA world, then the need for subscriptions would disappear; and while overall expenditure might rise for some research-intensive institutions, it might fall for others. But so long as we have a mixed economy, both subscriptions and APCs - for articles published by fully-OA journals and publishers, as well as OA articles in hybrid journals - will be needed. Hence the concerns that overall expenditure may increase, both for the UK as a whole and for individual institutions.

6.4. The Finch Group's 2013 review of progress in the implementation of its original recommendations[13] noted that there has already been much discussion about the likely potential for offsets between expenditure on APCs and subscriptions, at both UK and local institutional levels; and the issue is referred to again in David Willetts' letter to Dame Janet Finch in response to that review. Determining how such offsets might be achieved – and their scale – is likely to be complex, and is certainly beyond the scope of the framework we propose. But providing some reliable information as evidence to be considered during the discussions should be helpful to all parties.

6.5. We propose, therefore, that data should be gathered annually from a stratified sample of universities relating to their overall expenditure on articles and journals. We set out in Annex D how such a sample of universities might be created. The aim should be to work closely with those

[13] *A Review of Progress in Implementing the Recommendations of the Finch Report*, available at http://www.researchinfonet.org/wp-content/uploads/2013/02/Final-version.pdf

universities to ensure that data is gathered on as comprehensive and consistent a basis as possible, taking account of the practicalities and of the differences between universities and their systems.

6.6. We propose that the data to be gathered should cover so far as possible:

a. total expenditure on APCs, and the number of articles for which they have been paid

b. total expenditure on other publication charges such as page and colour charges, and the number of articles for which they have been paid

c. total expenditure on subscriptions to peer-reviewed scholarly journals, including both print and e-journals, and the number of journals subscribed to[14].

6.7. We recognise that gathering this information in a consistent format will depend on universities having the necessary systems in place to record the different transactions involved, and that it will be necessary to work with them to ensure consistency of practice across institutions in dealing with a range of issues including payments relating to membership schemes and deposit accounts; delays in invoicing, payments and publication; inadequacies in metadata; and so on..

6.8. We also considered whether further information should be gathered in these annual exercises, including information relating to administrative costs. We concluded, however, that it would be extremely difficult to gather data on administrative costs on a comprehensive and consistent basis; and interpreting such data would be subject to judgements about the extent and quality of the administrative services provided.

6.9. Ideally, however, and subject to an assessment of practicalities and the additional costs involved, we propose that consideration should be given to gathering data on

a. universities' expenditure from Wellcome Trust and RCUK block grants to meet the costs of APCs and other publication charges, and for other purposes; and their use of other funds to meet APCs; and

b. the overall numbers of articles published by authors affiliated to the university each year; and the numbers deposited in, or freely accessible via, their repositories (noting that 'accessible via' could relate to access provided through metadata links).

Embargo periods and subscriptions

6.10. A key concern from the perspective of many publishers and learned societies has been that short embargo periods bring with them the danger of reductions in the number of subscriptions to their journals, and thus risks to the journals' viability. There have been a number of attempts to assess the nature and extent to any dependency between embargo periods and potential risk to journals, making use for the most part of proxy data on half-lives of downloads. It is of course very difficult to assess the scale of risk in a rapidly-changing environment where past data may not provide a useful guide to future behaviour; and some commentators suggest that there is no risk at all.

---

[14] If at all possible, the data should include estimates of the number of non-OA articles in the journals subscribed to, although there may be practical difficulties in generating such estimates in a consistent fashion.

6.11. Given the complexity of the issues, we have concluded reluctantly that it is not possible to provide useful evidence on the basis of a simple data-gathering exercise. The issue is of undoubted importance, however, and further work is needed to clarify the nature of any relationship between embargoes and subscriptions. We recommend that any further work of this kind meets two important criteria: first, it should be based on evidence of actual rather than suggested changes in behaviour; and second, it should articulate a clear chain of causality, with sensitivity analyses associated with each element in the chain.

Market shares and ease of entry

6.12. From the supply side, questions relating to the overall structure of the market both in the UK and globally are clearly important from a number of different perspectives. We have therefore considered whether as part of our proposed framework it could be useful to monitor trends in the market shares of publishers grouped by category (commercial, not-for-profit, learned society), by size in terms of numbers of journals and articles published globally, and perhaps at journal level by OA/non-OA category (fully OA, hybrid, delayed OA, subscription-only). The overall shape at present is well-known, with a relatively small number of large and medium-sized publishers, and a long tail of much smaller publishers, some of which publish only one or two journals. But the rise of the new born-OA publishers has already had an impact on that shape; and other factors may come into play during – and in part as a result of – the transition to OA.

6.13. A broad indication of changes may be garnered from the data and indicators on accessibility which we outline in Section 3 above. But in order to generate a more detailed picture, we propose that data on numbers of titles (fully-OA, hybrid, delayed access and subscription-only) and of articles published in each category of journal should be gathered from the major publishers. An audit check should be undertaken on a sample of the journals, and indicators produced on the market shares of different categories of publisher.

6.14. A related issue is the ease with which new products and services can enter the market: one of the assumed benefits of OA is that it reduces the barriers to entry into the market for new players. Hence we considered whether it would be straightforward to detect trends in the entry of new players. We concluded that any data that could be readily collected was unlikely to be comprehensive, and would be difficult to interpret. Hence we do not propose that the framework should make any attempt to address the issue, at least for the near future[15].

Price signals and market effectiveness

6.15. The Finch Group noted that one of the arguments in favour of Gold OA is that it makes pricing more transparent to researchers. The effectiveness of the market in terms of price differentiation and

[15] Any exercise to assess barriers to entry would need, of course, to gather and analyse data on trends in the actual entry of new players

signals to researchers, universities and funders has been the subject of a recent study by Bjork and Solomon[16], and we may anticipate further attention being paid to this issue.

6.16. The extent to which researchers, their institutions and funders are influenced by the price of APCs when making their decisions on where to publish is an empirical one, and we propose that it should be addressed through a mixture of critical incident and conjoint analysis questions in the survey of researchers that we propose in paragraphs 7.2-7.5 below.

Learned societies

6.17. The potential impact on UK learned societies of the transition to OA was a matter of significant concern for the Finch Group, since many societies depend for a significant part of their overall income on the surpluses generated by their journals. The concern is that moves towards OA would bring significant risks to those surpluses, and thus to societies' ability to fund a wide range of their activities in support of the research community. Moreover, substantial proportions of societies' publishing revenues derive from overseas sales, which are thus available to contribute to their work in support of the UK and broader research base. Such concerns are widely shared, and some work related to the roles of and the risks to learned societies has been initiated and commissioned by HEFCE and by the ESRC.

6.18. We have considered how best to gather and analyse useful data on these issues as part of the framework. We propose that data on the overall income and expenditure – as well as the volumes of journal-related income and expenditure – should be gathered by surveying a sample of societies. The construction of the sample will have to be undertaken with care, and it will need to be sufficiently large to cover the various kinds of societies. Moreover, the survey data will require careful interpretation, to take account of factors such as the launch of new journals, new contracts with publishers and so on. We believe that such data should provide a valuable – albeit broad – overall picture in which individual societies will be able to locate their own position. It should also provide valuable evidence for more detailed studies.

---

[16] Bo-Christer Björk and David Solomon *Developing an Effective Market for Open Access Article Processing Charges*, 2014. Accessible at http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055910.pdf

Summary

- Data should be gathered annually from a stratified sample of universities on their expenditure on
  - APCs and on other publication charges, and the number of articles for which they have been paid
  - journal subscriptions, including both print and e-journals
- Ideally, information should also be gathered on universities' expenditure from Wellcome Trust and RCUK block grants to meet the costs of APCs and for other purposes; their use of other funds to meet APCs; and on the overall numbers of articles published, and the number deposited in the institutional repository.
- Data on the numbers of titles (fully-OA, hybrid, delayed access and subscription-only) and of articles published in each category of journal should be gathered from the major publishers. An audit check of numbers should be undertaken on a sample of the journals
- In the survey we propose in Section 7, authors should be asked questions to assess the extent to which they and their institutions are influenced by the price of APCs when making decisions on where to publish
- Data should be gathered from a sample of UK learned societies on their overall as well as their journal-related income and expenditure.

7. Quality of services to readers and authors

Author services

7.1. It is often argued that the quality of the services provided to authors by journals – many of them based in the UK – represents one of the key underpinnings for the success of the UK research community. Some commentators have suggested that the transition to OA might have an impact – for good or ill - on the quality of those services, with a possible consequent impact on the performance of UK researchers. The Government has commissioned successive comparative studies of UK levels of performance[17], and we assume that such studies will continue. But in the context of the transition to OA we believe that it will be important to monitor trends in authors' perceptions of the quality of service they receive from journals of different kinds.

7.2. We therefore propose that annual surveys should be undertaken of authors' perceptions and experiences in their interactions with journals. Our preference would be for each survey to make use of a random sample of UK researchers drawn from a reputable and comprehensive database. If that is impracticable on grounds of cost or data protection, then a stratified sample of publishers could be asked to send emails to a random sample of authors with a request to complete a survey instrument held on a single website.

7.3. The surveys should use a critical incident technique focusing on authors' most recent completed interaction with a journal (leading either to rejection or to acceptance and publication). In addition to basic information about themselves (position, institution, but not even necessarily name) and about the subject matter of the article, questions would relate to their levels of satisfaction with the service they have received in relation to that particular publication or submission. They would also be asked to name the journal and state why they chose it. Specific questions would also relate (but not be limited) to issues such as

a. whether the article had been submitted to other journals before it was finally accepted

b. whether it was revised or amended as a result of the peer review process before final acceptance, and the nature and extent of such revisions

c. authors' satisfaction with the levels of service at key stages of the publication process

d. the time elapsed between submission and a first editorial decision, and between notification of acceptance and publication.

7.4. We are aware, of course, of the limitations of such surveys. But we believe that a survey of this kind, conducted annually, would provide valuable information on trends.

7.5. We also suggest a further development of the survey, using conjoint analysis techniques to probe the factors which underlay respondents' choice of journal: speed of decision and of publication,

---

[17] The most recent was the *International Comparative Performance of the UK Research Base – 2013*, prepared by Elsevier for the Department of Business Innovation and Skills.: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263729/bis-13-1297-international-comparative-performance-of-the-UK-research-base-2013.pdf (viewed 31 January 2014).

status in the field (however authors might assess that), reach and breadth of audience, previous experience with the journal, compliance with funders' policies, need to pay an APC (and its level) etc. Such an analysis could provide valuable information about the drivers and the barriers affecting the pace and the extent of the shift to OA. The challenge, as in all conjoint analysis exercises, would be to achieve an appropriate balance between over-simple and over-complex approaches, while keeping costs acceptable.

Reader services

7.6. Services to readers[18] are critically important to the efficient conduct of research and to enhancing its impact. They have developed enormously in the past decade: search and navigation services; linking to other publications and to underlying datasets; semantic enrichment of various kinds; metrics (including 'alt-metrics') again of many different kinds; and reference management and sharing services. Such services are provided by a range of agencies including publishers (both OA and non-OA), aggregators and secondary publishers of different kinds, libraries, and many new intermediaries and service providers.

7.7. The development and take-up of these many different services is a highly-significant part of the scholarly communications landscape, and some of them are facilitated by OA. But we see no *necessary* relationship between the development and use of high-quality reader services and the transition to OA. Moreover, we can see no straightforward way of gathering useful data in relation to reader services. Hence we have concluded that, for the time being at least, they should not feature in the framework we propose.

Summary

- ❐ Annual surveys should be undertaken to assess authors' perceptions and experiences in relation to their most recent completed interaction with a journal
- ❐ We do not recommend at this stage a parallel survey of readers and users.

[18] The primary and critically-important service that publishers provide for readers, of course, is to publish and disseminate high-quality reports of the latest research.

8. Costs and governance

8.1. We are aware that, while we have tried to keep issues relating to costs and cost-effectiveness in mind in developing our proposals, implementation will require the deployment of significant resources. In terms of the overall budgets of research funders and universities, and even of the information-related budgets of libraries and publishers, the amounts are small, however, and we believe that they are well-justified. Our estimates of likely costs in terms of person-days required to undertake the work we propose are set out in Annex E.

8.2. In line with our principle of joint ownership and responsibility, it is important that the processes of gathering and analysing data, and generating indicators of the kind we propose, should be overseen by a body that includes representatives of the key stakeholder groups: universities, research funders, libraries, publishers and learned societies, as well as researchers themselves. Such a body is now being convened by Universities UK, and we believe that it should take on this oversight role.

8.3. Different agencies with access to relevant data, and the necessary expertise, should take responsibility for undertaking the work, operating in concert with others as necessary and appropriate. But it will be important that the UUK-convened body should give close attention to the work as it proceeds, provide a sounding board as the work progresses and issues arise, and take full responsibility for what is published and made available for others to use. It should also review the reception of the work, and suggestions for improvements and refinements to the proposals we set out here. As noted in Section 1, we recommend a light-touch review after the first annual exercise, and a more thoroughgoing review after the second.

Annex A

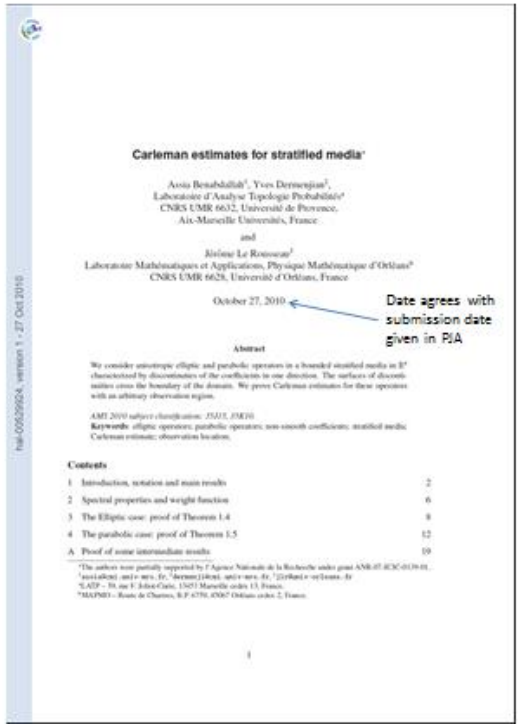
Membership of the Working Group

Mayur Amin (Elsevier, for the Publishers Association)
Simon Bell (British Library)
Laura Bellingan (Society of Biology)
Steven Hill (HEFCE)
Neil Jacobs (Jisc)
Michael Jubb (Research Information Network, Chair)
Robert Kiley (Wellcome Trust)
Valerie McCutcheon (University of Glasgow, for ARMA)
David Prosser (Research Libraries UK)
Wim van der Stelt (BioMedCentral, for the Publishers Association)
Mari Williams (BBSRC, for RLUK)

Annex B

Criteria for identifying versions of articles

Example 1: Preprint



Carleman estimates for stratified media*

Assia Benabdallah[1], Yves Dermenjian[2],
Laboratoire d'Analyse Topologie Probabilités[a]
CNRS UMR 6632, Université de Provence,
Aix-Marseille Universités, France

and

Jérôme Le Rousseau[3]
Laboratoire Mathématiques et Applications, Physique Mathématique d'Orléans[b]
CNRS UMR 6628, Université d'Orléans, France

October 27, 2010

Date agrees with submission date given in PIA

Abstract

We consider anisotropic elliptic and parabolic operators in a bounded stratified media in $\mathbb{R}^n$ characterized by discontinuities of the coefficients in one direction. The surfaces of discontinuities cross the boundary of the domain. We prove Carleman estimates for these operators with an arbitrary observation region.

AMS 2010 subject classification: 35J15, 35K10.

Keywords: elliptic operators; parabolic operators; non-smooth coefficients; stratified media; Carleman estimate; observation location.

Contents

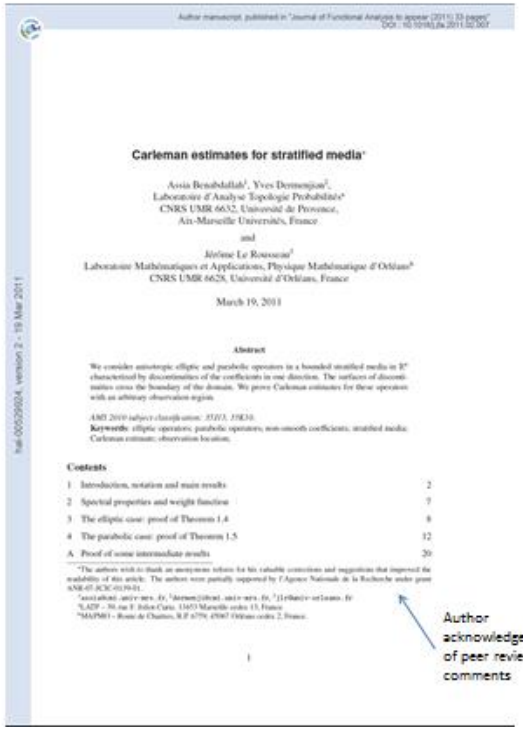
1 Introduction, notation and main results 2
2 Spectral properties and weight function 6
3 The Elliptic case: proof of Theorem 1.4 8
4 The parabolic case: proof of Theorem 1.5 12
A Proof of some intermediate results 19

*The authors were partially supported by l'Agence Nationale de la Recherche under grant ANR-07-JCJC-0139-01.

[1]assia@cmi.univ-mrs.fr, [2]dermen@cmi.univ-mrs.fr, [3]jrl@univ-orleans.fr
[a]LATP – 39, rue F. Joliot-Curie, 13453 Marseille cedex 13, France.
[b]MAPMO – Route de Chartres, B.P 6759, 45067 Orléans cedex 2, France.

hal-00529924, version 1 - 27 Oct 2010

Example 1: Accepted Author Manuscript



Author manuscript, published in "Journal of Functional Analysis to appear (2011) 33 pages"
DOI : 10.1016/j.jfa.2011.02.007

Carleman estimates for stratified media*

Assia Benabdallah[1], Yves Dermenjian[2],
Laboratoire d'Analyse Topologie Probabilités[a]
CNRS UMR 6632, Université de Provence,
Aix-Marseille Universités, France

and

Jérôme Le Rousseau[3]
Laboratoire Mathématiques et Applications, Physique Mathématique d'Orléans[b]
CNRS UMR 6628, Université d'Orléans, France

March 19, 2011

Abstract

We consider anisotropic elliptic and parabolic operators in a bounded stratified media in $\mathbb{R}^n$ characterized by discontinuities of the coefficients in one direction. The surfaces of discontinuities cross the boundary of the domain. We prove Carleman estimates for these operators with an arbitrary observation region.

AMS 2010 subject classification: 35J15, 35K10.

Keywords: elliptic operators; parabolic operators; non-smooth coefficients; stratified media; Carleman estimate; observation location.

Contents

1 Introduction, notation and main results 2
2 Spectral properties and weight function 7
3 The elliptic case: proof of Theorem 1.4 8
4 The parabolic case: proof of Theorem 1.5 12
A Proof of some intermediate results 20

*The authors wish to thank an anonymous referee for his valuable corrections and suggestions that improved the readability of this article. The authors were partially supported by l'Agence Nationale de la Recherche under grant ANR-07-JCJC-0139-01.

[1]assia@cmi.univ-mrs.fr, [2]dermen@cmi.univ-mrs.fr, [3]jrl@univ-orleans.fr
[a]LATP – 39, rue F. Joliot-Curie, 13453 Marseille cedex 13, France
[b]MAPMO – Route de Chartres, B.P 6759, 45067 Orléans cedex 2, France.

Author acknowledgement of peer review comments

hal-00529924, version 2 - 19 Mar 2011

Example 1: Published Journal Article

ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 260 (2011) 3645–3677
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Carleman estimates for stratified media ☆

Assia Benabdallah [a], Yves Dermenjian [a], Jérôme Le Rousseau [b,*]

[a] *Laboratoire d'Analyse Topologie Probabilités, CNRS UMR 6632, Université de Provence, Aix-Marseille Universités, 39, rue F. Joliot-Curie, 13453 Marseille cedex 13, France*
[b] *Laboratoire Mathématiques et Applications, Physique Mathématique d'Orléans, CNRS UMR 6628, Université d'Orléans, Route de Chartres, B.P. 6759, 45067 Orléans cedex 2, France*

Received 27 October 2010; accepted 4 February 2011
Available online 4 March 2011
Communicated by J. Coron

Abstract

We consider anisotropic elliptic and parabolic operators in a bounded stratified media in $\mathbb{R}^n$ characterized by discontinuities of the coefficients in one direction. The surfaces of discontinuities cross the boundary of the domain. We prove Carleman estimates for these operators with an arbitrary observation region.

© 2011 Elsevier Inc. All rights reserved.

Keywords: Elliptic operators; Parabolic operators; Non-smooth coefficients; Stratified media; Carleman estimate; Observation location

1. Introduction, notation and main results

Consider a bounded open set $\Omega \subset \mathbb{R}^n$. For a second-order elliptic operator, say $A = -\Delta_x$, Carleman estimates take the form[1]

$$s^3 \|e^{s\varphi} u\|^2_{L^2(\Omega)} + s \|e^{s\varphi} \nabla_x u\|^2_{L^2(\Omega)} \leq \|e^{s\varphi} Au\|^2_{L^2(\Omega)}, \quad u \in \mathcal{C}_0^\infty(\Omega),\ s \geq s_0.$$

☆ The authors wish to thank an anonymous referee for his valuable corrections and suggestions that improved the readability of this article. The authors were partially supported by l'Agence Nationale de la Recherche under grant ANR-07-JCJC-0139-01.

* Corresponding author.
E-mail address: jrl@univ-orleans.fr (J. Le Rousseau).

[1] $a \lesssim b$ stands for $a \leq Cb$ for some constant $C > 0$.

0022-1236/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jfa.2011.02.007

Example 2: Published Journal Article

Computational Statistics and Data Analysis 56 (2012) 3705–3713

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data

Patrick Borges[a,*], Josemar Rodrigues[b], Narayanaswamy Balakrishnan[c]

[a] Departamento de Estatística, Universidade Federal do Espírito Santo, Brazil
[b] Departamento de Estatística, Universidade Federal de São Carlos, Brazil
[c] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 15 July 2011
Received in revised form 15 October 2011
Accepted 15 October 2011
Available online 19 October 2011

Keywords:
Initiated cells
Cure rate models
IGPS distribution
Additional parameter ρ
Correlation structure

ABSTRACT

In this paper, we propose a new cure rate survival model, which extends the model of Rodrigues et al. (2011) by incorporating a structure of dependence between the initiated cells. To create the structure of the correlation between the initiated cells, we use an extension of the generalized power series distribution by including an additional parameter ρ (the inflated-parameter generalized power series (IGPS) distribution, studied by Kolev and Mendonça (2000)). It has a natural interpretation in terms of both a "zero-inflated" proportion and a correlation coefficient. In our approach, the number of initiated cells is assumed to follow the IGPS distribution. The IGPS distribution is a natural choice for modeling correlated count data that exhibit overdispersion. The primary advantage of this distributional assumption is that the correlation structure induced by the additional parameter ρ results in a natural characterization of the association between the initiated cells. Moreover, it provides a simple and realistic interpretation for the biological mechanism of the occurrence of the event of interest as it includes a process of destruction of tumor cells after an initial treatment or the capacity of an individual exposed to irradiation to repair initiated cells that result in cancer being induced. This means that what is recorded is only the undamaged portion of the original number of initiated cells not eliminated by the treatment or repaired by the repair system of an individual. Parameter estimation of the proposed model is then discussed through the maximum likelihood estimation procedure. Finally, we illustrate the usefulness of the proposed model by applying it to a real cutaneous melanoma data.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Cure rate survival modelling plays an important role in reliability and survival analysis. It pertains to survival studies wherein a proportion of the subjects might not be susceptible to the event of interest due to different competing causes. These models have found important applications in such diverse areas as biomedical studies, finance, criminology, demography, manufacturing and industrial reliability. For instance, in biomedical data, an event of interest can be a patient's death, which can occur due to different competing causes or a tumor recurrence that may arise due to a number of metastasis-component tumor cells that are left active after an initial treatment of the individual. A metastasis-component tumor cell is a tumor cell which has the potential for metastasizing; see Yakovlev (1994), Yakovlev et al. (1993), Yakovlev and Tsodikov (1996), and Ibrahim et al. (2001).

* Correspondence to: Departamento de Estatística, Universidade Federal do Espírito Santo, Av. Fernando Ferrari 514, Goiabeiras, CEP 29075-910, Vitória, Espírito Santo, Brazil.
E-mail address: patrick@ufes.edu.br (P. Borges).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.10.015

Example 2: Accepted Author Manuscript

Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data

Patrick Borges[a], Josemar Rodrigues[b] and Narayanaswamy Balakrishnan[c]

[a] Departamento de Estatística, Universidade Federal do Espírito Santo, Brazil
[b] Departamento de Estatística, Universidade Federal de São Carlos, Brazil
[c] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

Abstract

In this paper, we propose a new cure rate survival model, which extends the model of Rodrigues et al. (2011) by incorporating a dependence structure between the initiated cells. To create the correlation structure between the initiated cells, we use an extension of the generalized power series distribution by including an additional parameter ρ (inflated-parameter generalized power series (IGPS) distribution, studied by Kolev (2000)). It has a natural interpretation in terms of both "zero-inflated" proportion and correlation coefficient. In our approach, the number of initiated cells is assumed to follow the IGPS distribution. The IGPS distribution is a natural choice for modeling correlated count data that exhibit overdispersion. The primary advantage of this distributional assumption is that the correlation structure induced by the additional parameter ρ results in a natural characterization of the association between the initiated cells. Moreover, it provides a simple and realistic interpretation for the biological mechanism of the occurrence of the event of interest as it includes a destructive process of tumor cells after an initial treatment or the capacity of an individual exposed to irradiation to repair initiated cells that results in cancer infraction. This means that, what is recorded is only the undamaged portion of the original number of initiated cells not eliminated by the treatment or repaired by the repair system of an individual. Parameter estimation of the proposed model is then discussed through the maximum likelihood estimation procedure. Finally, we illustrate the usefulness of the proposed model by applying it to a real cutaneous melanoma data.

Key words: Initiated cells, Cure rate models, IGPS distribution, Additional parameter ρ, Correlation structure.

1. Introduction

Cure rate survival modelling plays an important role in reliability and survival analysis. It pertains to survival studies wherein a proportion of the subjects might not be susceptible to the event of interest due to different competing causes. These models have found important applications in such diverse

*Corresponding author: Patrick Borges. Departamento de Estatística, Universidade Federal do Espírito Santo, Av. Fernando Ferrari 514, Goiabeiras, CEP 29075-910, Vitória, Espírito Santo, Brazil. Email: patrick@ufes.edu.br

Example 2: Accepted Author Manuscript

6. Concluding remarks

In this paper, we extend the models of Rodrigues et al. (2011) by proposing a model for survival data in the presence of latent dependent competing causes (tumour cells) and a cure fraction. We assume a IGPS distribution for the number of initiated cells and a Weibull model for the lifetimes, and obtain the so called CDGPS model. The CDGPS model incorporates into the analysis a biological dependence between the tumor cells. The advantage of this assumption is that we can measure the interdependence between the cells in an initiated tissue developing into a malignant tumor. The estimation procedure for the parameters of the proposed model is achieved through the maximum likelihood method. The practical relevance and applicability of the model are then demonstrated with a real data involving cutaneous melanoma. The proposed model, in addition to offering better interpretation to the underlying biological mechanism, offers better fit than the other commonly used cure rate models. The proposed CDGPS models will be quite helpful in the understanding of the biological process for a variety of infections, irradiation carcinogenesis and cancer chemoprevention experiments.

Acknowledgments. The authors express their sincere thanks to the Associate Editor and anonymous

referee for making some useful comments on an earlier version of this manuscript, which led to this improved version.

Author acknowledgement of peer review comments

References

Aarset, M.V., 1985. The null distribution for a test of constant versus bathtub failure rate. Scandinavian Journal of Statistics 12, 55–68.

Banerjee, S., Carlin, B.P., 2003. Parametric spatial cure rate model for interval-censored time-to-relapse data. Biometrics 59, 269–275.

Example 4: Published Journal Article

Geophysical and Astrophysical Fluid Dynamics
Vol. 106, No. 3, June 2012, 305–319

Taylor & Francis

Slow quasigeostrophic unstable modes of a lens vortex in a continuously stratified flow

HAI YEN NGUYEN, BACH LIEN HUA*,
RICHARD SCHOPP and XAVIER CARTON

Laboratoire de Physique des Oceans, IFREMER, BP 70,
29280 Plouzané, France

(Received 18 March 2011; in final form 9 August 2011; first published online 6 October 2011)

This work addresses the linear dynamics underlying the formation of density interfaces at the periphery of energetic vortices, well outside the vortex core, both in the radial and axial directions. We compute numerically the unstable modes of an anticyclonic Gaussian vortex lens in a continuously stratified rotating fluid. The most unstable mode is a slow mode, associated with a critical layer instability located at the vortex periphery. Although the most unstable disturbance has a characteristic vertical scale which is comparable to the vortex height, interestingly, the critical levels of the successively fastest growing modes are closely spaced at intervals along the axial direction that are much smaller than the vortex height.

Keywords: Quasigeostrophic model; Vortices; Stratified flow; Critical level; Flow instability

1. Introduction

Persistent layering, defined here as the vertical stacking of sharp interfaces in temperature and density, has been recently evidenced at the periphery of energetic oceanic vortices using seismic imaging of the water column (Biescas *et al.* 2008, Menesguen *et al.* 2009a, Papenberg *et al.* 2010, Quentel *et al.* 2010). Typical values of the observed vertical scales involved in the layering is in the range of a few meters up to about 100 m. What is novel is the documentation that such long-lived layering is abundant not only at the lateral periphery of vortices, as emphasized in early observational work (Armi *et al.* 1989), but is even stronger both aloft and below the eddy core, at the vertical periphery of the vortex. Another striking observation is the lateral spatial coherence of the layering which can extend over distances comparable to the vortex horizontal size for several tens of kilometers. These two characteristic properties of the layering, its localization at the periphery of oceanic intra-thermocline vortices and its large lateral coherence, are the focus of this study. Here, we present adiabatic dynamical arguments for sharp interfaces formation that depart from previously invoked rationales for layering that invoke, either different diffusive

*Corresponding author. Email: hua@ifremer.fr

ISSN 0309-1929 print/ISSN 1029-0419 online © 2012 Taylor & Francis
http://dx.doi.org/10.1080/03091929.2011.620568

Downloaded by [IFREMER - Centre De Brest] at 04:59 23 May 2012

Example 4: Accepted Author Manuscript

Clear labelling of manuscript stage

Bibliographic details of subsequent PIA version

Geophysical & Astrophysical Fluid Dynamics
June 2012, Volume 106, Issue 3, Pages 305-319
http://dx.doi.org/10.1080/03091929.2011.620568
© 2012 Taylor & Francis

Archimer
http://archimer.ifremer.fr

The original publication is available at http://www.tandf.co.uk/journals/

Slow quasigeostrophic unstable modes of a lens vortex in a continuously stratified flow

Hai Yen Nguyen, Bach Lien Hua*, Richard Schopp and Xavier Carton

IFREMER, Laboratoire de Physique des Oceans, BP 70, 29280 Plouzané, France

* Corresponding author : Bach Lien Hua, email address : Bach.Lien.Hua@ifremer.fr

Abstract:

This work addresses the linear dynamics underlying the formation of density interfaces at the periphery of energetic vortices, well outside the vortex core, both in the radial and axial directions. We compute numerically the unstable modes of an anticyclonic Gaussian vortex lens in a continuously stratified rotating fluid. The most unstable mode is a slow mode, associated with a critical layer instability located at the vortex periphery. Although the most unstable disturbance has a characteristic vertical scale which is comparable to the vortex height, interestingly, the critical levels of the successively fastest growing modes are closely spaced at intervals along the axial direction that are much smaller than the vortex height.

Keywords: Quasigeostrophic model ; Vortices ; Stratified flow ; Critical level ; Flow instability

1. Introduction

Persistent layering, defined here as the vertical stacking of sharp interfaces in temperature and density, has been recently evidenced at the periphery of energetic oceanic vortices using seismic imaging of the water column (Biescas *et al.* 2008, Menesguen *et al.* 2009a, Papenberg *et al.* 2010, Quentel *et al.* 2010). Typical values of the observed vertical scales involved in the layering is in the range of a few meters up to about 100 meters. What is novel is the documentation that such long-lived layering is abundant not only at the lateral periphery of vortices, as emphasized in early observational work (Armi *et al.* 1989), but is even stronger both aloft and below the eddy core, at the vertical periphery of the vortex. Another striking observation is the lateral spatial coherence of the layering which can extend over distances comparable to the vortex horizontal size for several tens of kilometers. These two characteristic properties of the layering, its localization at the periphery of oceanic intra-thermocline vortices and its large lateral coherence, are the focus of the present study. Here, we present adiabatic dynamical arguments for sharp interfaces formation that depart from previously invoked rationales for layering that invoke, either different diffusive properties of temperature and salt in the ocean(e.g. Ruddick and Richards 2003), or which require an inertial instability of the flow (e.g. Menesguen *et al.* 2009b).
We build on the study by Benilov (2003) which dealt with the possibility of a slow destabilization of oceanic vortices through critical level instabilities at the lateral periphery of vortices based on an asymptotic analysis of a thin vortex in a two-layer configuration. In the present paper we explore numerically how his results on critical level instabilities can generalize to the case of an intra-thermocline vortex lens in a continuously stratified fluid.
The above cited work is part of an abundant literature on oceanic vortices stability which started with Gent and McWilliams (1986), Flierl (1988), and Helfrich and Send (1988), and were motivated

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher authenticated version is available on the publisher Web site.

Example 4: Accepted Author Manuscript

Clear labelling of manuscript stage

Bibliographic details of subsequent PIA version

Geophysical & Astrophysical Fluid Dynamics
June 2012, Volume 106, Issue 3, Pages 305-319
http://dx.doi.org/10.1080/03091929.2011.620568
© 2012 Taylor & Francis

Archimer
http://archimer.ifremer.fr

The original publication is available at http://www.tandf.co.uk/journals/

Slow quasigeostrophic unstable modes of a lens vortex in a continuously stratified flow

Hai Yen Nguyen, Bach Lien Hua*, Richard Schopp and Xavier Carton

IFREMER, Laboratoire de Physique des Oceans, BP 70, 29280 Plouzané, France

* Corresponding author : Bach Lien Hua, email address : Bach.Lien.Hua@ifremer.fr

Abstract:

This work addresses the linear dynamics underlying the formation of density interfaces at the periphery of energetic vortices, well outside the vortex core, both in the radial and axial directions. We compute numerically the unstable modes of an anticyclonic Gaussian vortex lens in a continuously stratified rotating fluid. The most unstable mode is a slow mode, associated with a critical layer instability located at the vortex periphery. Although the most unstable disturbance has a characteristic vertical scale which is comparable to the vortex height, interestingly, the critical levels of the successively fastest growing modes are closely spaced at intervals along the axial direction that are much smaller than the vortex height.

Keywords: Quasigeostrophic model ; Vortices ; Stratified flow ; Critical level ; Flow instability

1. Introduction

Persistent layering, defined here as the vertical stacking of sharp interfaces in temperature and density, has been recently evidenced at the periphery of energetic oceanic vortices using seismic imaging of the water column (Biescas *et al.* 2008, Menesguen *et al.* 2009a, Papenberg *et al.* 2010, Quentel *et al.* 2010). Typical values of the observed vertical scales involved in the layering is in the range of a few meters up to about 100 meters. What is novel is the documentation that such long-lived layering is abundant not only at the lateral periphery of vortices, as emphasized in early observational work (Armi *et al.* 1989), but is even stronger both aloft and below the eddy core, at the vertical periphery of the vortex. Another striking observation is the lateral spatial coherence of the layering which can extend over distances comparable to the vortex horizontal size for several tens of kilometers. These two characteristic properties of the layering, its localization at the periphery of oceanic intra-thermocline vortices and its large lateral coherence, are the focus of the present study. Here, we present adiabatic dynamical arguments for sharp interfaces formation that depart from previously invoked rationales for layering that invoke, either different diffusive properties of temperature and salt in the ocean(e.g. Ruddick and Richards 2003), or which require an inertial instability of the flow (e.g. Menesguen *et al.* 2009b).
We build on the study by Benilov (2003) which dealt with the possibility of a slow destabilization of oceanic vortices through critical level instabilities at the lateral periphery of vortices based on an asymptotic analysis of a thin vortex in a two-layer configuration. In the present paper we explore numerically how his results on critical level instabilities can generalize to the case of an intra-thermocline vortex lens in a continuously stratified fluid.
The above cited work is part of an abundant literature on oceanic vortices stability which started with Gent and McWilliams (1986), Flierl (1988), and Helfrich and Send (1988), and were motivated

1

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher authenticated version is available on the publisher Web site.

Annex C

Moving cohort samples for articles

Cohort sampling approach for assessing OA accessibility													
Date of publication/ testing	**Jan-12**	**Jun-12**	**Jan-13**	**Jun-13**	**Jan/Feb 14**	**Jun/Jul 14**	**Jan/Feb 15**	**Jun/Jul 15**	**Jan/Feb 16**	**Jun/Jul 16**	**Jan/Feb 17**	**Jun/Jul 17**	**Jan/Feb 18**
6 month intervals													
Cohort 1 (A-E)													
published	A	B	C	D	E								
tested					A25 months, B19 months, C13 months, D7 months, E1 month	B25, C19, D13, E7	C25, D19, E13	D25, E19	E25				
Cohort 2 (F)													
published						F							
tested						F1	F7	F13	F19	F25			
Cohort 3 (G)													
published							G						
tested							G1	G7	G13	G19	G25		
Cohort 4 (H)													
published								H					
tested								H1	H7	H13	H19	H25	
Cohort 5 (I)													
published									I				
tested									I1	I7	I13	I19	I25
Cohort 6 (J)													
published										J			
tested										J1	J7	J13	J19
12 month intervals													
Cohort 1 (A-D)													
published	A		B	C	D								
tested					A25 months, B13 months,C7 months, D1 month		B25,D13		D25				
Cohort 2 (E-F)													
published						E	F						
tested							E7, F1		F13		F25		
Cohort 3 (G-H)													
published								G	H				
tested									G7, H1		H13		
Cohort 4 (I-J)													
published										I	J		
tested											I7, J1		

Annex D

Stratified sampling for universities

Our proposed sampling approach for identifying universities which would be asked to provide information on expenditure on APCs and on subscriptions would involve the creation of a sample of 18 HEIs, after

a. identifying three groups of institutions based on the number of articles and proceedings published in 2012, with the size of each group increasing as the number of publications falls; and randomly selecting 6 from each group (so that the chance of selection is higher for the most highly-productive institutions)
b. assessing that sample and then modifying it to ensure that it includes
    i. a reasonable geographical spread across the UK, including at least one institution from Wales, and two from Scotland
    ii. at least one specialist institution
    iii. an overall spread of publications by subject (as in the four REF panels) that is broadly in line with the proportions for the UK as a whole.

Annex E

Costs

We have sought throughout our discussions to avoid the risk of specifying tasks that are over-ambitious and too costly. We cannot be sure about the costs of the key tasks set out in our recommendations, but we have drawn on evidence from others who have undertaken similar work as to the number of days required. Our estimates are set out below.

1. identifying publishers and journals making use of the four different routes to OA	15 days
2. creating samples of published articles, undertaking automated queries to identify accessible versions, and manual checking of those versions	60 days
3. gathering and audit of data from major publishers on 'supply' of OA options, and on numbers of journals and articles published	20 days
4. gathering and analysing expenditure data for 18 universities	20 days
5. survey of income and expenditure from learned societies	15 days
6. author survey	20 days, depending on extent and use of external address lists
7. oversight and management of data gathering and analysis	15 days

The overall total here is 165 days, with a fair proportion of the work to be undertaken at junior level, but a need also for significant input at more senior level.