Bibliometric analysis of the Research performance of NIOO-KNAW 2012

Ed Noyons
Center for Science and Technology Studies (CWTS)
Leiden University
The Netherlands

CWTS report July 2012

Executive summary

This study reports on the performance of NIOO-KNAW in terms of scientific impact as measured by citations of NIOO-KNAW scholarly publication output. It is a quantitative analysis using the oeuvre as far as covered by the Web of Science (WoS).

> bibliometric indicators can be valuable to assess the current state of scholarly output

The data in this report measure the impact of the publication oeuvre of a research entity as measured by citations. They do not assess the *quality*, in whatever way defined. It helps to understand the focus and impact of the research unit under study and also the trend therein. For a content quality assessment, peer review or a similar evaluation is needed. Bibliometric measures can be supportive to that process.

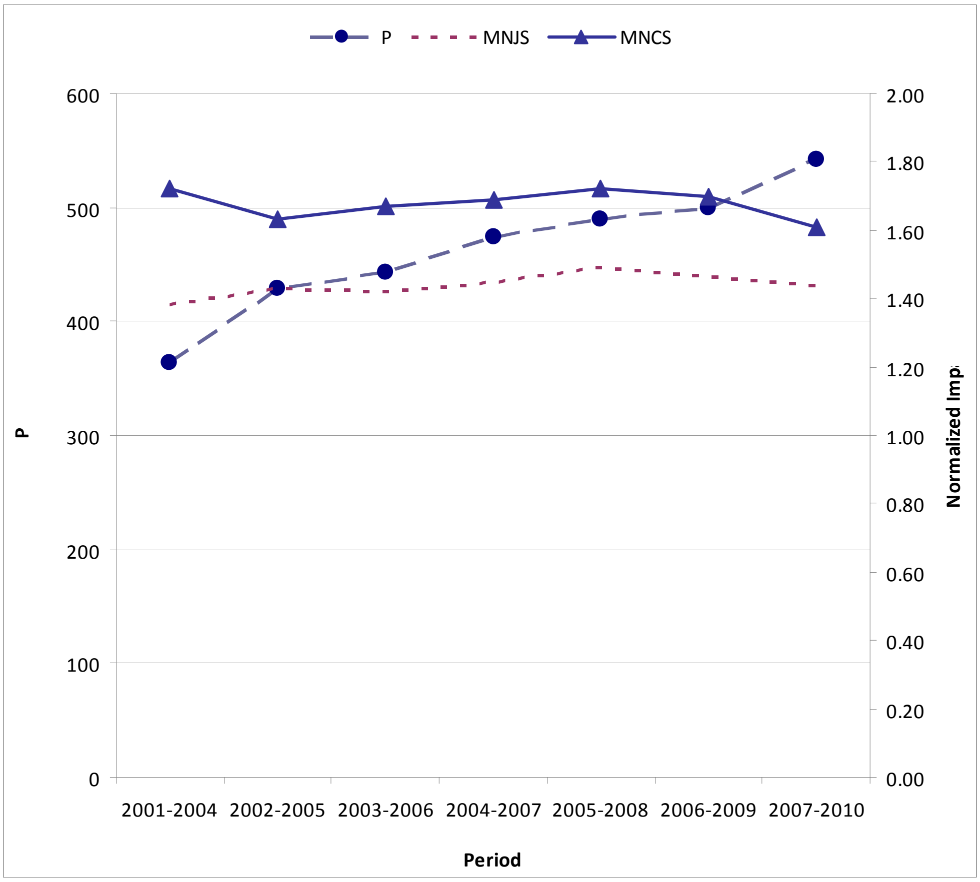
The data in this report cover the period 2001-2010/11. We emphasize that the performance of the individual departments of NIOO-KNAW should only be viewed from within their own contexts (e.g., a trend analysis) because it may lead to inadequate conclusions.

> The impact [...] is way above world average

The entire volume of output is over 1,100 (WoS) papers in ten years (2001-2010) and grows steadily during the studied period. This WoS oeuvre of NIOO-KNAW approximately covers 83% of its entire scholarly output. (MNCS, world average = 1, in diagram below). Moreover, NIOO-KNAW gets their papers published in high impact journals (MNJS, field average=1). The slight decrease of (MNCS) impact in the most recent period co-occurs with a slight decrease of impact of journals chosen (MNJS). We often see this happening when the production increases substantially. The increase of production may have been established by submitting to lower impact journals or by an increased activity of young researchers whose work has not yet reached a 'mature' impact.

> [...] comparison of impact between the departments [...] not recommended

Figure 1: NIOO-KNAW production, normalized impact and impact of journals used, 2001-2010



At the level of departments, we found that all four NIOO-KNAW departments have a substantial output volume over the years. Moreover, the impact in all cases is well above world average, albeit at different levels. Furthermore, we found that each department has its own characteristic profile.

The department *Animal Ecology* shows a steep increase regarding production volume during the studied period 2001-2010. During the 10 years the amount of papers was more than doubled. Moreover the impact is way above world average and is for an important part achieved from the national collaboration and in the core fields of its research (*Ecology*, *Ornithology* and *Zoology*). In the most recent years the impact is still at a high level but somewhat below the earlier period. The steep increase of output volume may have had an effect on the choice of journals as well as on the impact.

The *Aquatic Ecology* department seems to be disadvantaged by the current normalization of impact based on the WoS classification in general and more in particular by a recent publication increase and strategy shift towards general ecological journals in the category *Ecology*. Still this department has an impact well above world average as well as a steep increase of production in the most recent period.

The department of *Microbial Ecology* is characterized by the recovery in terms of impact from 2003 onwards, with a steady increase of production. Since 2004 this department publishes their papers in journals with a higher impact each year. Since 2005 we found a stable high impact level until present.

Finally, *Terrestrial Ecology* can be characterized as the most stable department. In the ten years of our analyses the impact remains almost unchanged at a high level. This is the only department where the production remains at a similar level (at least from 2002 onwards).

Table of Contents

1 Introduction

This study reports on the performance of NIOO-KNAW in terms of scientific impact as measured by citations to NIOO-KNAW scholarly publication output. It is a quantitative analysis using the oeuvre as far as covered by the Web of Science (WoS), a bibliographic database covering international peer refereed journals in all fields of science. This database discloses bibliographic data (including citations) of all articles in journals and is processed by CWTS for bibliometric analysis.

It should be mentioned in advance that quantitative analyses tend to lead to spreadsheet management. Assessment of progress and quality based on figures are popular but conceal a lot of valuable information. We realize that the results of a bibliometric analysis create an appealing report to assess the achievements of a research entity. It should be noted, however, that the data in this report measure the impact of the publication oeuvre of a research entity as measured by citations. No more and no less. They do not assess the *quality*, in whatever way defined. In that sense the results have their limitations. Still, within their own context bibliometric measured can be used as valuable indicators for the current state of the scholarly output. It helps to understand the focus and impact of the research unit under study as well as the trend therein.

The data in this report cover the period 2001-2010/11. As CWTS has recently improved the quality of its indicators (c.f. Waltman 2011a and 2011b), we cover a longer period than necessary to provide sufficient data for a trend analysis. We argue that the performance of the individual departments of NIOO-KNAW can only be compared within its own context, not with the performance of the others.

2 Bibliometric indicators

At CWTS, we normally calculate our indicators based on our in-house version of the Web of Science (WoS) database of Thomson Reuters. WoS is a bibliographic database that covers the publications of about 12,000 journals in the sciences, the social sciences, and the arts and humanities. Each journal in WoS is assigned to one or more subject categories. These subject categories can be interpreted as scientific fields. There are about 250 subject categories in WoS. Some examples are Astronomy & Astrophysics, Economics, Philosophy, and Surgery. Publications in multidisciplinary journals such as Nature, Proceedings of the National Academy of Sciences, and Science were individually allocated to subject fields on the basis of their references.. Each publication in WoS has a document type. The most frequently occurring document types are article, book review, correction, editorial material, letter, meeting abstract, news item, and review. In the calculation of bibliometric indicators, we only take into account publications of the document types article, letter, and review. Publications of other document types usually do not make a significant scientific contribution. We note that our in-house version of the WoS database includes a number of improvements over the original WoS database. Most importantly, our database uses a more advanced citation matching algorithm and an extensive system for address unification. Our database also supports a hierarchically organized field classification system on top of the WoS subject categories. We note that at the moment conference proceedings are not covered by our database. In the future, however, our database will also include conference proceedings.

To determine the appropriateness of our indicators for assessing a particular research entity, we often look at the internal WoS coverage of the entity. The internal WoS coverage of an entity is defined as the proportion of the references in its oeuvre that points to publications (also) covered by WoS. The lower the internal WoS coverage of an entity's oeuvre, the more careful one should be in the interpretation of our indicators.

The rest of this section provides an in-depth discussion of the bibliometric indicators that we use in this report.

Overview of the bibliometric indicators discussed in this section.

Indicator	*Dimension*	*Definition*
P	Output	Total number of publications of a research group.
Int_cov	Output	Internal coverage. Proxy of oeuvre being covered by Web of Science. Measured by the proportion of cited references in the oeuvre linking to other WoS publications
MCS	Impact	Average number of citations of the publications of a research group (self-citations not included).
MNCS	Impact	Average normalized number of citations of the publications of a research group (self-citations not included).
$PP_{top\ 10\%}$	Impact	Proportion publications of a research group belonging to the top 10% most frequently cited publications in their field (self-citations not included).
MNJS	Journal impact	Average normalized citation score of the journals in which a research group has published (self-citations not included).

2.1 Indicators of output

To measure the total publication output of a research group, we use a very simple indicator. This is the number of publications indicator, denoted by **P**. This indicator is calculated by counting the total number of publications of a research unit.

2.2 Indicators of Impact

A number of indicators are available for measuring the average scientific impact of the publications of a research group. These indicators are all based on the idea of counting the number of times the publications of a research group have been cited. Citations can be counted using either a fixed-length citation window or a variable-length citation window. In the case of a fixed-length citation window, only citations received within a fixed time period (e.g., three years) after the appearance of a publication are counted. In the case of a variable-length citation window, all citations received by a publication up to a fixed point in time are counted, which means that older publications have a longer citation window than more recent publications. An advantage of a variable-length window over a fixed-length window is that a variable-length window usually yields higher citation counts, which is be expected to lead to more reliable impact measurements. A disadvantage of a variable-length window is that citation counts of older and more recent publications cannot be directly compared with each other. Using a variable-length window, older publications on average have higher citation counts than more recent publications, which makes direct comparisons impossible. This difficulty does not occur with a fixed-length window. At CWTS, we mostly work with a variable-length window, where citations are counted up to and including the most recent year fully covered by our database. In trend analyses, however, we usually use a fixed-length window. This ensures that different publication years are treated in the same way as much as possible. Furthermore, in the calculation of our impact indicators, we only take into account publications with a citation window of at least one full year. For instance, if our database covers publications until the end of 2011, this means that publications from 2011 are not taken into account, while publications from 2010 are.

In the calculation of our impact indicators, we disregard author self citations. We classify a citation as an author self citation if the citing publication and the cited publication have at least one author name (i.e., last name and initials) in common. We disregard self citations because they have a somewhat different nature than ordinary citations. Many self citations are given for good reasons, in particular to indicate how different publications of a researcher build on each other. However, sometimes self citations serve mainly as a mechanism for self promotion rather than as a mechanism for indicating relevant related work. This is why we consider it preferable to exclude self citations from the calculation of our impact indicators. By disregarding self citations, the sensitivity of our impact indicators to manipulation is reduced. Disregarding self citations means that our impact indicators focus on measuring the impact of the work of a researcher on other members of the scientific community. The impact of the work of a researcher on his own future work is ignored.

Our most straightforward impact indicator is the mean citation score indicator, denoted by MCS. This indicator simply equals the average number of citations of the publications of a research group. Only citations within the relevant citation window are counted, and author self citations are excluded. Also, only citations to publications of the document types: article, letter, and review are taken into account. In the calculation of the average number of citations per publication, articles and reviews have a weight of one while letters have a weight of 0.25.

A major shortcoming of the MCS indicator is that it cannot be used to make comparisons between scientific fields, because different fields have very different citation characteristics. For instance, using a three-year fixed-length citation window, the average number of citations of a publication of the document type article equals 2.0 in mathematics and 19.6 in cell biology. So it clearly makes no sense to make comparisons between these two fields using the MCS indicator. Furthermore, when a variable-length citation window is used, the MCS indicator also cannot be used to make comparisons between publications of different ages. In the case of a variable-length citation window, the MCS indicator favors older publications over more recent ones because older publications tend to have higher citation counts.

Our mean normalized citation score indicator, denoted by MNCS, provides a more sophisticated alternative to the MCS indicator. The MNCS indicator is similar to the MCS indicator except that it performs a normalization that aims to correct for differences in citation characteristics between publications from different scientific fields, between publications of different ages (in the case of a variable-length citation window), and between publications of different document types (i.e., article, letter, and review[1]). To calculate the MNCS indicator for a research group, we first calculate the normalized citation score of each publication of the group. The normalized citation score of a publication equals the ratio of the actual and the expected number of citations of the publication, where the expected number of citations is defined as the average number of citations of all publications in WoS that belong to the same field and that have the same publication year and the same document type. The field (or the fields) to which a publication belongs is determined by the WoS subject categories of the journal in which the publication has appeared. The MNCS indicator is obtained

[1] We note that the distinction between the different document types is sometimes based on somewhat arbitrary criteria. This is especially the case for the distinction between the document types *article* and *review*. One of the main criteria used by WoS to distinguish between these two document types is the number of references of a publication. In general, a publication with fewer than 100 references is classified as *article* while a publication with at least 100 references is classified as *review*. It is clear that this criterion does not yield a very accurate distinction between ordinary articles and review articles.

by averaging the normalized citation scores of all publications of a research group. As in the case of the MCS indicator, letters have a weight of 0.25 in the calculation of the average while articles and reviews have a weight of one. If a research group has an MNCS indicator of one, this means that on average the actual number of citations of the publications of the group equals the expected number of citations. In other words, on average the publications of the group have been cited equally frequently as publications that are similar in terms of field, publication year, and document type. An MNCS indicator of, for instance, two means that on average the publications of a group have been cited twice as frequently as would be expected based on their field, publication year, and document type. We refer to Waltman, Van Eck, Van Leeuwen, Visser, and Van Raan (2011) for more details on the MNCS indicator.

To illustrate the calculation of the MNCS indicator, we consider a hypothetical research group that has only five publications. Table 1 provides some bibliometric data for these five publications. For each publication, the table shows the scientific field to which the publication belongs, the year in which the publication appeared, and the actual and the expected number of citations of the publication. (For the moment, the last column of the table can be ignored.) The five publications are all of the document type article. Citations have been counted using a variable-length citation window. As can be seen in the table, publications 1 and 2 have the same expected number of citations. This is because these two publications belong to the same field and have the same publication year and the same document type. Publication 5 also belongs to the same field and has the same document type. However, this publication has a more recent publication year, and it therefore has a smaller expected number of citations. It can further be seen that publications 3 and 4 have the same publication year and the same document type. The fact that publication 4 has a larger expected number of citations than publication 3 indicates that publication 4 belongs to a field with a higher citation density than the field in which publication 3 was published. The MNCS indicator equals the average of the ratios of actual and expected citation scores of the five publications. Based on Table 1, we obtain

$$\text{MNCS} = \frac{1}{5}\left(\frac{7}{6.13} + \frac{37}{6.13} + \frac{4}{5.66} + \frac{23}{9.10} + \frac{0}{1.80}\right) = 2.08$$

Hence, on average the publications of our hypothetical research group have been cited more than twice as frequently as would be expected based on their field, publication year, and document type.

Table 1: Bibliometric data for the publications of a hypothetical research group.

Publication	*Field*	*Year*	*Actual citations*	*Expected citations*	*Top 10% threshold*
1	Surgery	2007	7	6.13	15
2	Surgery	2007	37	6.13	15
3	Clinical neurology	2008	4	5.66	13
4	Hematology	2008	23	9.10	21
5	Surgery	2009	0	1.80	5

In addition to the MNCS indicator, we have another important impact indicator. This is the *proportion top 10% publications indicator*, denoted by $PP_{top\ 10\%}$. For each publication of a research group, this indicator determines whether based on its number of citations the publication belongs to the top 10% of all WoS publications in the same field (i.e., the same WoS subject category) and the same publication year and of the same document type. The $PP_{top\ 10\%}$ indicator equals the proportion of the publications of a research group that belong to the top 10%. As for the MCS and MNCS indicators, letters are given less weight than articles and reviews in the calculation of the $PP_{top\ 10\%}$ indicator. If a research group has a $PP_{top\ 10\%}$ indicator of 10%, this means that the actual number of top 10% publications of the group equals the expected number. A $PP_{top\ 10\%}$ indicator of, for instance, 20% means that a group has twice as many top 10% publications as expected. Of course, the choice to focus on top 10% publications is somewhat arbitrary. Instead of the $PP_{top\ 10\%}$ indicator, we can also calculate for instance a $PP_{top\ 1\%}$, $PP_{top\ 5\%}$, or $PP_{top\ 20\%}$ indicator. In this study, however, we use the $PP_{top\ 10\%}$ indicator. On the one hand this indicator has a clear focus on high impact publications, while the indicator is more stable than for instance the $PP_{top\ 1\%}$ indicator.

To illustrate the calculation of the $PP_{top\ 10\%}$ indicator, we use the same example as we did for the MNCS indicator. Table 1 shows the bibliometric data for the five publications of the hypothetical research group that we consider. The last column of the table indicates for each publication the minimum number of citations needed to belong to the top 10% of all publications in the same field and the same publication year and of the same document type.[2] Of the five publications, there are two (i.e., publications 2 and 4) whose number of citations is above the top 10% threshold. These two publications are top 10% publications. It follows that the $PP_{top\ 10\%}$ indicator equals

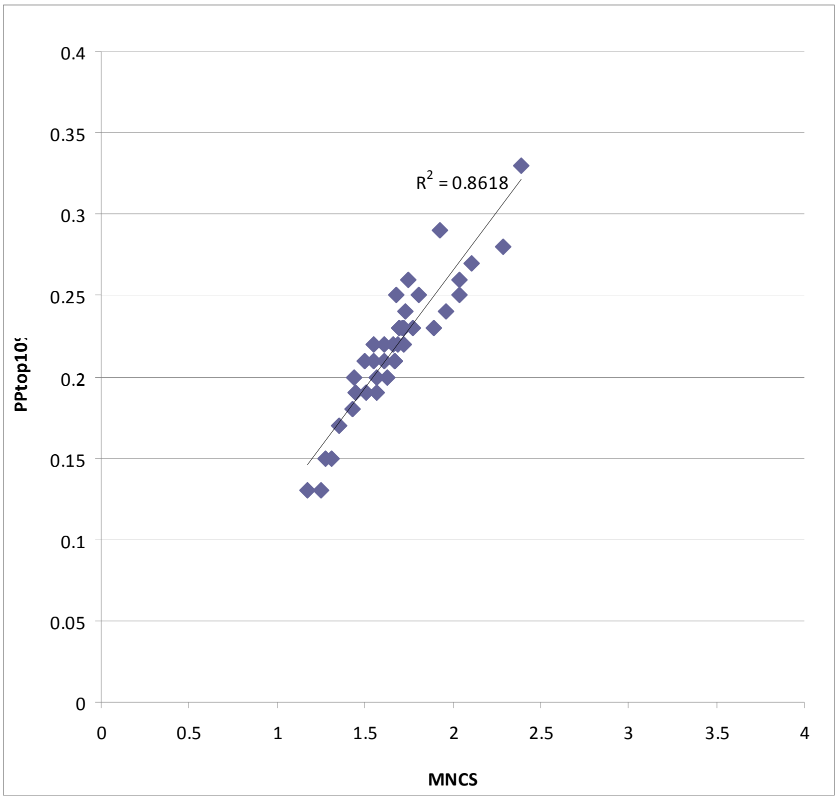
$$PP_{top\ 10\%} = \frac{2}{5} = 0.4 = 40\%$$

In other words, top 10% publications are four times overrepresented in the set of publications of our hypothetical research group.

To assess the impact of the publications of a research group, our general recommendation is to rely on a combination of the MNCS indicator and the $PP_{top\ 10\%}$ indicator. The MCS indicator does not correct for field differences and should therefore be used only for comparisons of groups that are active in the same field. An important weakness of the MNCS indicator is its strong sensitivity to publications with a very large number of citations. If a research group has one very highly cited publication, this is usually sufficient for a high score on the MNCS indicator, even if the other publications of the group have received only a small number of citations. Because of this, the MNCS indicator may sometimes overestimate significantly the actual scientific impact of the publications of a research group. The $PP_{top\ 10\%}$ indicator is much less sensitive to publications with a very large number of citations, and it therefore does not suffer from the same problem as the MNCS indicator. A disadvantage of the $PP_{top\ 10\%}$ indicator is the artificial dichotomy it creates between publications that belong to the top 10% and publications that do not belong to the top 10%. A publication whose number of citations is just below the top 10% threshold does not contribute to the $PP_{top\ 10\%}$ indicator, while a publication with one or two additional citations does contribute to the indicator. Because the MNCS indicator and the $PP_{top\ 10\%}$ indicator have more or less opposite strengths and weaknesses, the indicators are strongly complementary to each other, which is why we recommend taking

[2] If the number of citations of a publication is exactly equal to the top 10% threshold, the publication is partly classified as a top 10% publication and partly classified as a non-top-10% publication. This is done in order to ensure that for each combination of a field, a publication year, and a document type we end up with exactly 10% top 10% publications.

both into account when assessing the impact of a research group's publications. It should be noted that the two indicators correlate almost perfectly as can be seen in the diagram below in which we plotted all MNCS scores we calculated (both for entire period as for year blocks) for the four departments against the PPtop10% scores.



All MNCS scores against all PPtop10% scores reported for NIOO-KNAW and its four departments

It is important to emphasize that the correction for field differences that is performed by the MNCS and $PP_{top\ 10\%}$ indicators is only a partial correction. As already mentioned, the field definitions on which these indicators rely are based on the WoS subject categories. It is clear that, unlike these subject categories, fields in reality do not have well-defined boundaries. The boundaries of fields tend to be fuzzy, fields may be partly overlapping, and fields may consist of multiple subfields that each have their own characteristics. From the point of view of citation analysis, the most important shortcoming of the WoS subject categories seems to be their heterogeneity in terms of citation characteristics. Many subject categories consist of research areas that differ substantially in their density of citations. For instance, within a single subject category, the average number of citations per publication may be 50% larger in one research area than in another. The MNCS and $PP_{top\ 10\%}$ indicators do not correct for this within-subject-category heterogeneity. This can be a problem especially when using these indicators at lower levels of aggregation, for instance at the level of departments as evaluated in this report. At these levels, within-subject-category heterogeneity may significantly reduce the accuracy of the impact measurements provided by the MNCS and $PP_{top\ 10\%}$ indicators. For this reason we stress that the results for the departments should not be compared among each other. By looking at the trend we do provide a picture of the development of the impact but to compare the impact results among the departments, we know that the citation densities differ too much so that some departments' impact may be affected by a normalization as applied on the basis of these subject categories.

2.3 Indicators of journal impact

In addition to the average scientific impact of the publications of a research group, it may also be of interest to measure the average scientific impact of the journals in which a research group has published. In general, high-impact journals may be expected to have stricter quality criteria and a more rigorous peer review system than low-impact journals. Publishing a scientific work in a high-impact journal may therefore be seen as an indication of the quality of the work.

We use the mean normalized journal score indicator, denoted by MNJS, to measure the impact of the journals in which a research group has published. To calculate the MNJS indicator for a research group, we first calculate the normalized journal score of each publication of the group. The normalized journal score of a publication equals the ratio of on the one hand the average number of citations of all publications published in the same journal and on the other hand the average number of citations of all publications published in the same field (i.e., the same WoS subject category). Only publications in the same year and of the same document type are considered. The MNJS indicator is obtained by averaging the normalized journal scores of all publications of a research group. As for the impact indicators discussed in Section 2.2, letters are given less weight than articles and reviews in the calculation of the average. The

MNJS indicator is closely related to the MNCS indicator. The only difference is that instead of the actual number of citations of a publication the MNJS indicator uses the average number of citations of all publications published in a particular journal. The interpretation of the MNJS indicator is analogous to the interpretation of the MNCS indicator. If a research group has an MNJS indicator of one, this means that on average the group has published in journals that are cited equally frequently as would be expected based on their field. An MNJS indicator of, for instance, two means that on average a group has published in journals that are cited twice as frequently as would be expected based on their field.

In practice, journal impact factors reported in Thomson Reuters' Journal Citation Reports are often used in research evaluations. Impact factors have the advantage of being easily available and widely known. The use of impact factors is similar to the use of the MNJS indicator in the sense that in both cases publications are assessed based on the journal in which they have appeared. However, compared with the MNJS indicator, impact factors have the important disadvantage that they do not correct for differences in citation characteristics between scientific fields. Because of this disadvantage, impact factors should not be used to make comparisons between fields. The MNJS indicator, on the other hand, does correct for field differences (albeit with some limitations; see the discussion at the end of Section 2.2). When between-field comparisons need to be made, the use of the MNJS indicator can therefore be expected to yield significantly more accurate journal impact measurements than the use of impact factors.

2.4 Analyses of cognitive orientation: research profiles

Research profiles are based on the scientific fields in which a group publishes, with scientific field defined by Thomson Reuters' categorization of journals. We give two indicators: the number of publications in each scientific field and the average impact in each of these fields. These indicators are important in comparing a group's performance in its core research fields compared with more peripheral fields. Indicators are reported for the entire period 2001-2010/2011.

2.5 Indicators of scientific collaboration: scientific cooperation profiles

The indicators of scientific collaboration are based on an analysis of all addresses in papers published by a group. We first identified all papers authored by scientists from NIOO-KNAW only. To these papers we assigned the collaboration type *''**No collaboration**'*. With respect to the remaining papers we established (on the basis of the addresses) whether authors participated from other groups within the Netherlands (‘***National***’), and finally whether scientists are involved from other groups outside the Netherlands (collaboration type ‘***International***’). If a paper by a group is the result of collaboration with *both* another *Dutch* group and a group *outside* the Netherlands, it is marked with collaboration type ***international***.

The purpose of this indicator is to show how frequently a group has co-published papers with other groups, and how the impact of papers resulting from national or international collaboration compares to the impact of publications authored by scientists from a single research group. This analysis was conducted for the period 2001-2010/2011

2.6 Basic elements of bibliometric analysis

All the above indicators are important in a bibliometric analysis as they relate to different aspects of publication and citation characteristics. Generally, we consider ***MNCS,*** *in combination with* ***PP$_{top\ 10\%}$*** as the most important indicators. These indicators relate the measured impact of a research group or institute to a worldwide, field-specific reference value, by both comparing with the averages in the fields as well as the position in the actual distribution of impact over publications per field. They thus constitute a powerful set of internationally standardized impact indicators. This indicator enables us to rapidly assess whether the performance of a research institute/group or institute is significantly far below (indicator value < 0.5), below (indicator value 0.5 - 0.8), about (0.8 - 1.2), above (1.2 - 2.0), or far above (>2.0) the international impact standard of the field.

We would like to emphasize that the meaning of the numerical value of the indicator is related to the aggregation level of the entity under study. The higher the aggregation level, the larger the volume in publications and the more difficult it is to have an average impact significantly above the international level. At the 'meso-level' (e.g., a large institute, or faculty, about 500 – 1,000 publications per year), a ***MNCS*** value above 1.2 means that the institute's impact as a whole is significantly above (Western-) world average. The institute can be considered as a scientifically strong organization, with a high probability to find very good to excellent groups. Therefore, it is important to split up large institutes into smaller groups. Only this allows a more precise assessment of research performance. Otherwise, excellent work will be 'hidden' within the bulk of a large institute or faculty.

In this study we present the bibliometric results over a ten/ eleven years' period, namely the period 2001 – 2010/11. The impact related to the publications produced in the NIOO-KNAW and its departments in this entire period is calculated as follows: for publications from each of the publication years (2001 - 2010), citations are counted up to and including 2011. For example, a six year citation window is used for papers published in 2005, and a three year citation window for papers published in 2008. We excluded 2011 as a publication year, since impact measurement of the last year's output is statistically unreliable. Furthermore, we weighted *letters* and their impact as only one quarter of a publication and its impact, to prevent distortion of the results by a single highly cited *letter*. In the P indicator, *letters* are weighted by 0.25.

2.7 Stability intervals

The stability of an indicator relates to the sensitivity of the indicator to changes in the underlying set of publications. An indicator has a low stability if it is highly sensitive to changes in the set of publications based on which it is calculated. An indicator has a high stability if it is relatively insensitive to such changes. For instance, if a research group has one very highly cited publication and a number of poorly cited publications, the MNCS indicator for this group will be quite unstable. This is because the value of the MNCS indicator depends strongly on whether the group's highly cited publication is included in the calculation of the indicator or not. A research group whose publications all have similar citation scores will have a very stable MNCS indicator. In general, the larger the number of publications of a research group, the more stable the indicators calculated for the group. We note that the notion of stability is closely related to the notion of reliability discussed in Section 2.2. Citation analysis has a high reliability if the indicators that are used are stable. If the indicators are unstable, the reliability of citation analysis is low.

To determine the stability of an indicator, we use a so-called stability interval. Stability intervals are similar to confidence intervals, but they have a somewhat different interpretation. A stability interval indicates a range of values of an indicator that are likely to be observed if the underlying set of publications changes. For instance, the MNCS indicator may be equal to 1.50 for a particular research group, with a stability interval from 1.25 to 1.90. This means that the true value of the MNCS indicator equals 1.50 for this group, but that changes in the set of publications of the group may relatively easily lead to MNCS values in the range from 1.25 to 1.90. Clearly, the larger the stability interval of an indicator, the lower the stability of the indicator.

We construct our stability intervals as follows. Given a set of n publications, suppose that we want to construct a stability interval for the MNCS indicator. We then randomly draw a large number of samples (e.g., 1000 samples) from our set of publications. Each sample is drawn with replacement, which means that a publication may occur multiple times in the same sample. The size of each sample is n, which is the same as the size of the original set of publications. For each sample, we calculate the value of the MNCS indicator. This yields a distribution of sample MNCS values. We derive our stability interval for the MNCS indicator from this distribution. We are usually interested in a 95% stability interval. To obtain such an interval, we take the 2.5th and the 97.5th percentile of the distribution of sample MNCS values. These percentiles serve as the lower and the upper bound of our stability interval. We note that in the statistical literature our procedure for constructing stability intervals is known as bootstrapping (Efron & Tibshirani, 1993).

3 Data collection

The data collection for the NIOO-KNAW units was established in the following way. NIOO-KNAW provided CWTS with the bibliographic data of their publications (2001-present). This data was matched against the CWTS bibliometric database using key bibliographic information and checked by hand for false positives, i.e., publications accidentally assigned as a NIOO-KNAW publication. The bibliographic data was provided by NIOO-KNAW per department.

In the analysis we use data from 2001 onwards to enable more valuable trend analyses. Moreover, in the previous evaluation study, the indicators CWTS used were modified slightly so that a comparison with those results is no longer valid. By providing results for the period 2001 onwards in the present evaluation a comparison with earlier years is possible.

4 Results

In this section, we will discuss the results for the entire NIOO-KNAW and each of its departments. It should be mentioned that the results for a department should not be interpreted in comparison to the others. Each set of results stands on its own because the contexts (i.e., the basis of normalization) within which the impact of each is measured differs too much.

4.1 NIOO-KNAW 2001-2010/11

For each entity in our analysis we present the following results. We give an overview of basis indicators for the entire period as well as for the 4 year publication blocks. Furthermore, we provide a collaboration profile regarding production and impact using three different types. Finally, we provide an overview of production as well as impact distributed over the subject fields used in the study. In some case we will use additional results to further illustrate a point we wish to make regarding the results.

The NIOO-KNAW overall has a very good profile. The internal coverage is high throughout, indicating that over 83% of its scientific output is covered by the Web of Science. The production volume is large enough to do a bibliometric analysis but we cannot say anything about productivity because no input information is available. NIOO-KNAW has a high impact as measured by MNCS and PPtop10%. The MNCS overall is almost 70% above world average and has remained steady between 1.6 and 1.7 over the assessment period. The stability interval (blue area around the scores per year block in Figure 2) shows that the impact does not increase or drop significantly. The PPtop10% shows a similar high impact twice the expected level of 0.1. The impact of journals in which NIOO-KNAW gets its papers published, is at a high level of 40% above field average. The percentage of self-citations around 28% is normal.
Although not significant, we do see slight decrease of both impact and journal impact. This often co-occurs with an increased production. The analyses on the department level will reveal more on this.

Table 2: NIOO-KNAW output and impact indicators overall and trend (2001-2010/11)

Period	*P*	*Internal coverage*	*TCS*	*MCS*	*MNCS*	*PP top10%*	*MNJS*	*% self-cits*
2001-2010	*1130.50*	*0.83*	*9841*	*8.70*	*1.67*	*0.21*	*1.43*	*28*
2001-2004	363.50	0.79	3097	8.52	1.72	0.23	1.38	28
2002-2005	428.25	0.80	3850	8.99	1.63	0.20	1.43	28
2003-2006	442.25	0.81	4167	9.42	1.67	0.21	1.42	28
2004-2007	474.50	0.82	4872	10.27	1.69	0.22	1.45	26
2005-2008	490.00	0.83	5193	10.60	1.72	0.22	1.49	27
2006-2009	500.00	0.85	4885	9.77	1.70	0.23	1.46	28
2007-2010	541.75	0.86	4414	8.15	1.61	0.21	1.44	28

Figure 2: Normalized Impact (MNCS) of NIOO-KNAW 2001-2010/11

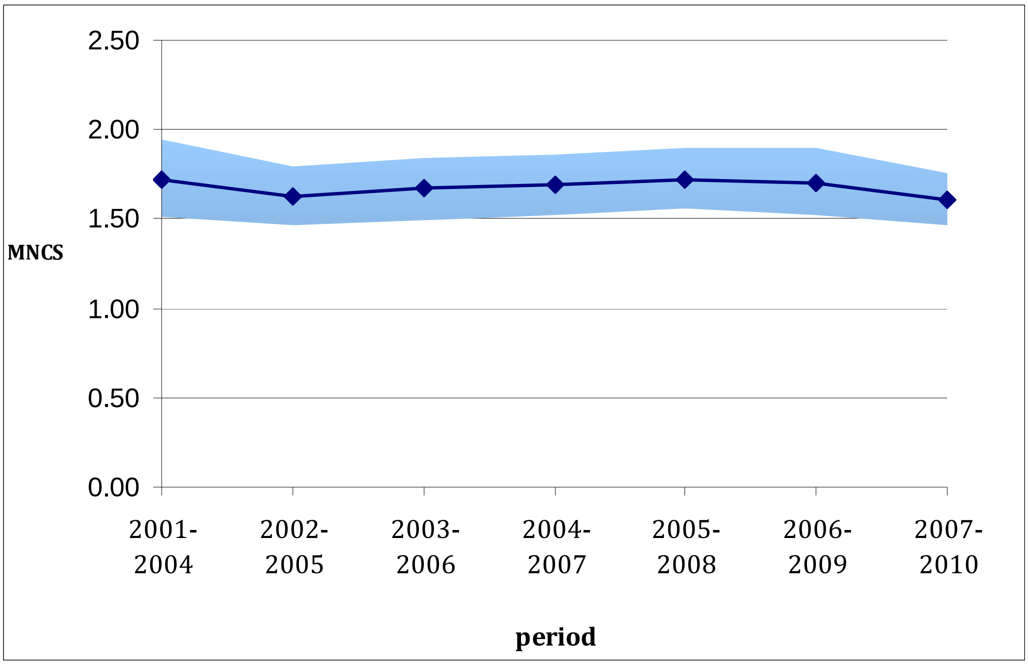
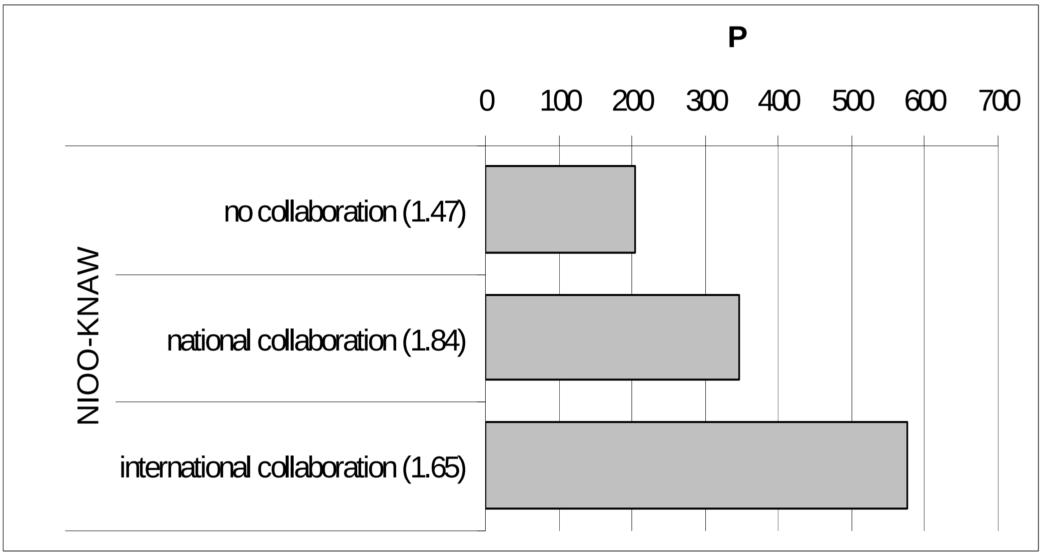
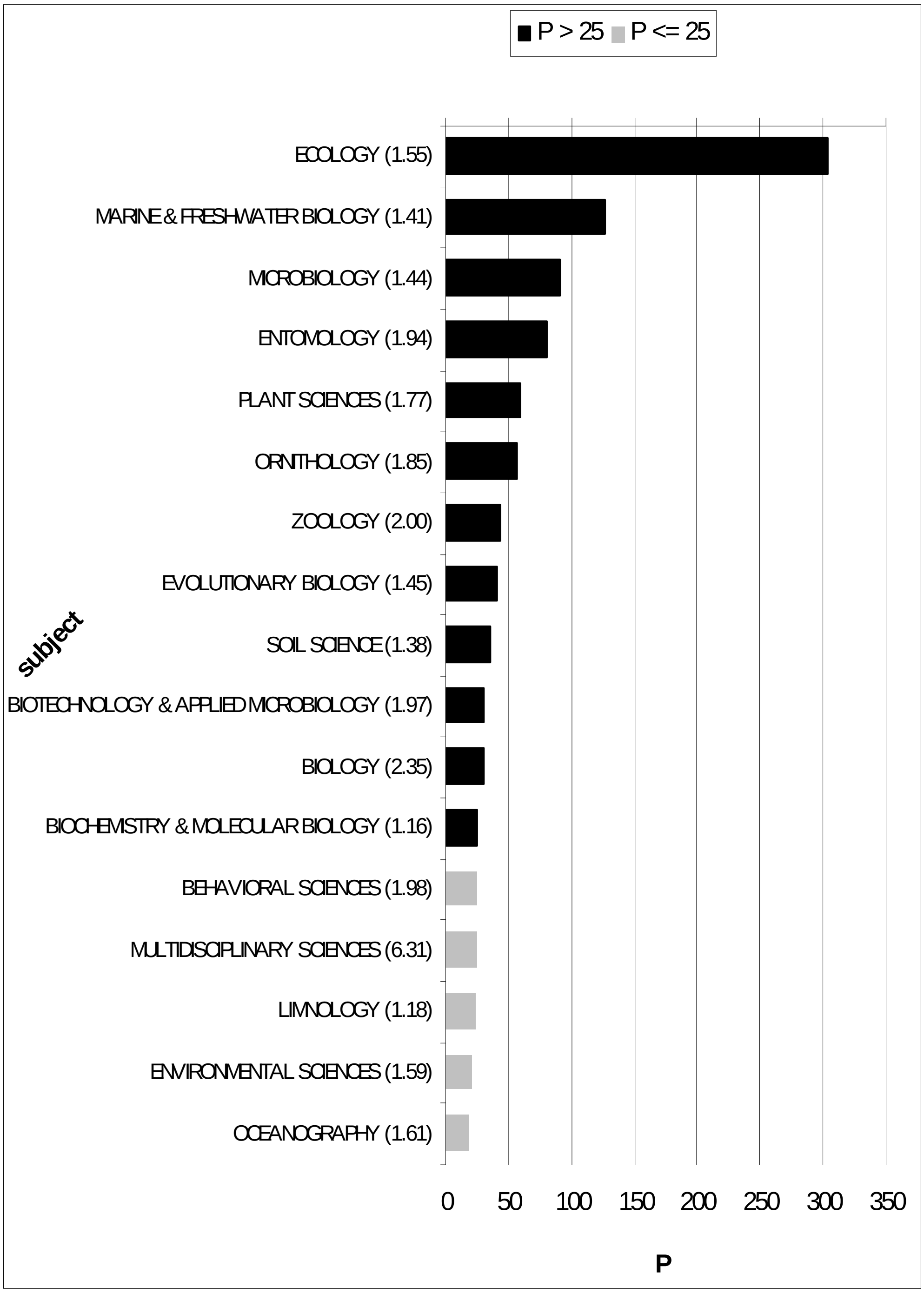


Figure 3: NIOO-KNAW collaboration output and impact (MNCS, between parentheses) 2001-2010/11



The collaboration profile (Figure 3) has a preference for international collaboration. The impact (between parentheses) is highest for output involving national collaboration.

Figure 4: NIOO-KNAW research profile and (between parentheses MNCS) impact (2001-2010/11)



If we look at the research profile of NIOO-KNAW (Figure 4) we see the obvious subject category *Ecology* on top with an impact of 50% above world average. But also the subject categories referring to the four departments *Marine & Fresh* *water Biology* (Aquatic Ecology), *Microbiology* (Microbial Ecology), *Entomology, Plant sciences* (Terrestrial Ecology) and *Ornithology, Zoology* (Animal Ecology) have a high impact overall.

4.2 NIOO-KNAW 2001-2010/11 by department

4.2.1 Animal Ecology

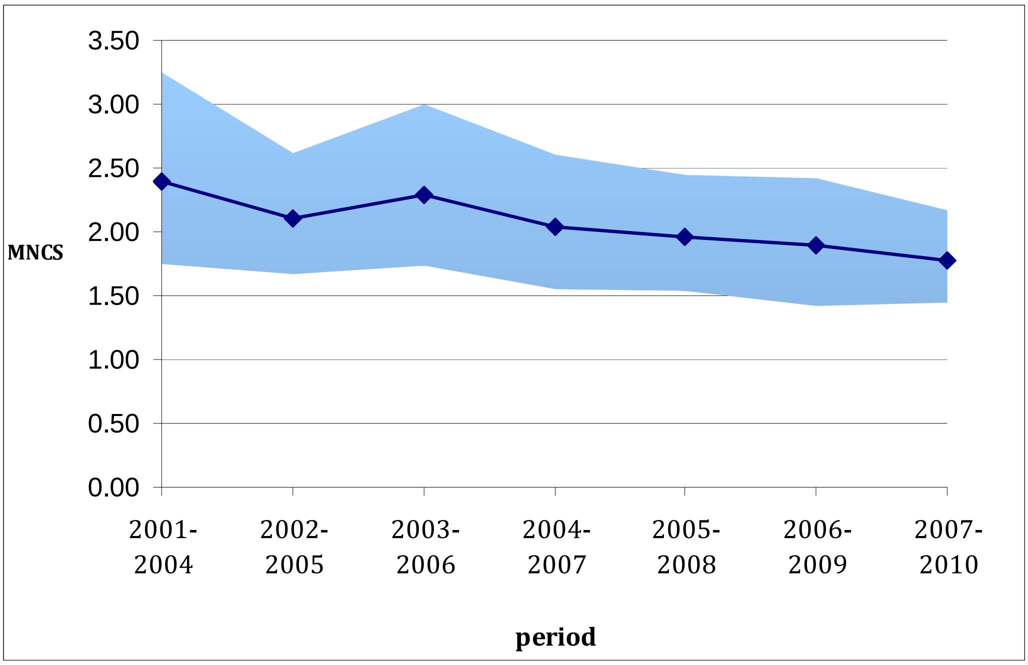
The NIOO-department Animal Ecology has a very good performance over the entire period. The almost 300 publications in 10 years have an impact (MNCS) of around or just below twice the world average and more than twice the expected value of PPtop10%. They get their papers published in journals with an impact of almost 70% above field average. Only in the most recent period Animal Ecology appears to choose journals with a slightly lower impact, still way above field average.

Table 3: Output and impact indicators overall and trend for department Animal Ecology (2001-2010/11)

period	*P*	*Internal coverage*	*TCS*	*MCS*	*MNCS*	*PP top10%*	*MNJS*	*% self cits*
Animal Ecology								
2001-2010	*288.00*	*0.82*	*2787*	*9.68*	*2.04*	*0.26*	*1.64*	*24*
2001-2004	70.25	0.76	742	10.56	2.39	0.33	1.70	25
2002-2005	89.25	0.78	978	10.96	2.11	0.27	1.74	23
2003-2006	98.25	0.8	1232	12.54	2.29	0.28	1.71	24
2004-2007	120.25	0.81	1438	11.96	2.04	0.25	1.73	23
2005-2008	135.50	0.82	1589	11.73	1.96	0.24	1.78	22
2006-2009	138.25	0.83	1414	10.23	1.89	0.23	1.57	24
2007-2010	152.50	0.84	1228	8.05	1.77	0.23	1.48	24

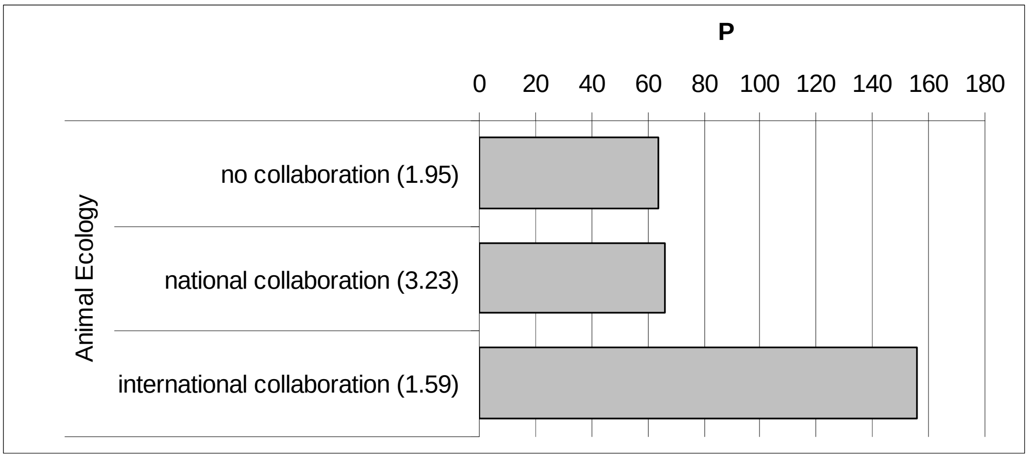
The exceptional high impact of around 2.4 at the start of the period studied drops to around 2 but remains within the stability interval (c.f. Figure 5). The production more than doubles during this period which usually has a negative on the impact (MNCS) as well as on the impact of the journals in which researchers publish their work. We see the latter happening from 2005 onwards.

Figure 5: Development of normalized impact (MNCS) of Department of Animal Ecology, 2001-2010/11 within stability intervals



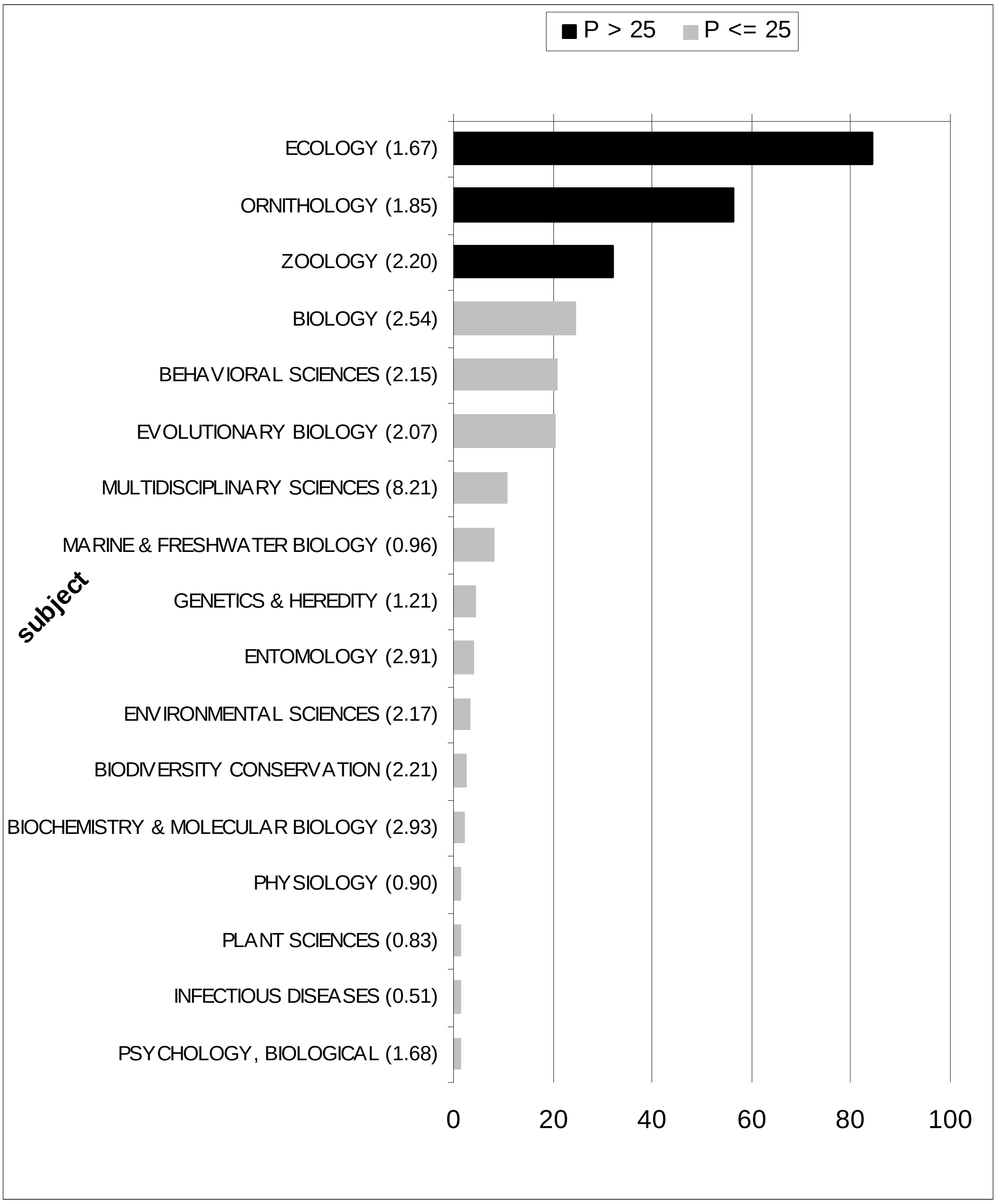
The collaboration profile has a preference for international collaboration but a remarkable impact measured for national collaboration. Obviously this is a particular and very successful collaboration.

Figure 6: Collaboration output and impact (MNCS, between parentheses) for department Animal Ecology (2001-2010/11)



The research profile of Animal Ecology has a clear focus on *Ecology*, *Ornithology* and *Zoology* with a high impact in each subject category.

Figure 7: Research profile for department Animal Ecology (2001-2010/11): output and impact (MNCS, between parentheses)



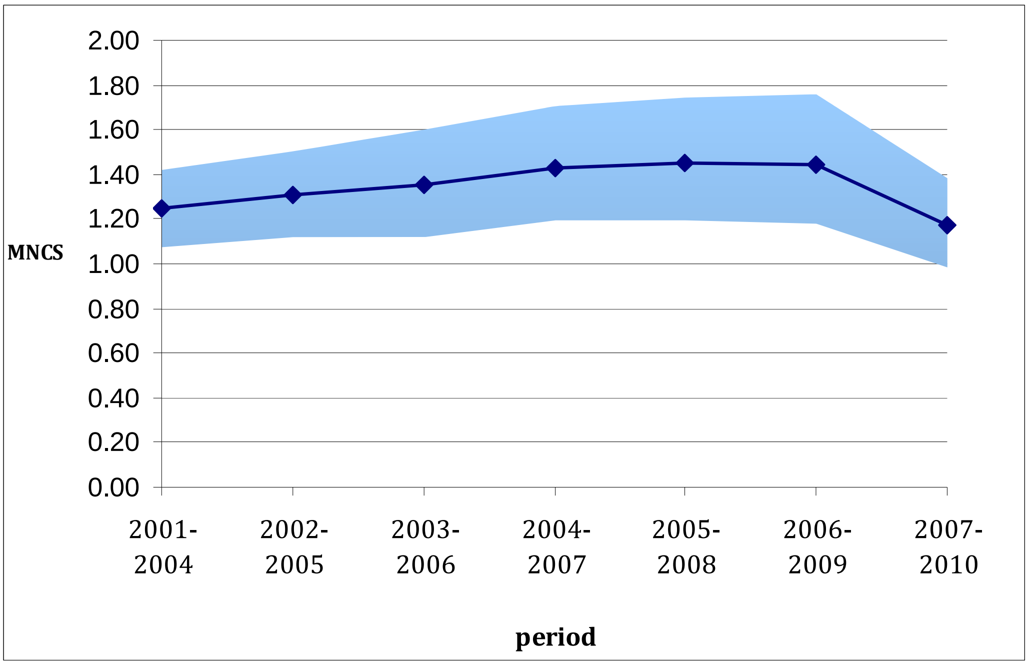
4.2.2 Aquatic Ecology

The NIOO-KNAW department Aquatic Ecology has a good performance over the entire period with more than 300 papers in 10 years, an MNCS around 30% above world average, a PPto10% fluctuating between 0.13 and 0.20 and gets its papers published in journals with an impact 20-30% above field average. In the recent years this has declined somewhat. Together with a steep increase of output, this may be the reason that the impact, as measured by MNCS and PPtop10%, is lower in the most recent years. It should be noted, however, that the decreasing impact remains within the margins of the stability interval. An interesting finding is that the proportion of self-citations appears to drop with 10%. This is the only department in which this happens. Researchers in this department appear to import increasingly knowledge from outside.

Table 4: Output and impact indicators overall and trend for department Aquatic Ecology (2001-2010/11)

period	*P*	*Internal coverage*	*TCS*	*MCS*	*MNCS*	*PP top10%*	*MNJS*	*% self cits*
Aquatic Ecology								
2001-2010	*306.00*	*0.78*	*1985*	*6.49*	*1.28*	*0.15*	*1.23*	*28*
2001-2004	124.00	0.74	656	5.29	1.25	0.13	1.29	35
2002-2005	136.00	0.75	886	6.51	1.31	0.15	1.31	32
2003-2006	131.00	0.77	947	7.23	1.35	0.17	1.31	30
2004-2007	121.00	0.79	1018	8.41	1.43	0.18	1.33	27
2005-2008	113.00	0.81	1031	9.12	1.45	0.19	1.25	23
2006-2009	113.00	0.83	960	8.50	1.44	0.20	1.23	23
2007-2010	130.00	0.82	784	6.03	1.17	0.13	1.14	25

Figure 8: Development of normalized impact (MNCS) of Department of Aquatic Ecology, 2001-2010/11 within stability intervals



It is clear that the impact as measured by MNCS and PPtop10% is somewhat lower than that of NIOO-KNAW and some of the other departments. We emphasize, however, that these figures should not be compared as such. In particular, we made an additional analysis on the output of Aquatic Ecology and found that the output does not properly match the subject categories in the Web of Science. A more fine grained classification would be required to normalize the impact of this department to world average.

As we can read the research profile, Aquatic Ecology acts on the interface of *Marine & Fresh Water Biology* and *Ecology* but seems to have changed the preference for *Marine & Fresh Water Biology* to *Ecology* journals. The situation in 2007-2010 is listed in Table 5. It shows that Aquatic Ecology is not yet able to reach the 'required' field average impact of *Ecology* in this period. The impact in the most recent period in *Marine & Fresh Water Biology* is still at a very high level.

Figure 9: Research profile for Department Aquatic Ecology (2001-2010/11): output and impact (MNCS, between parentheses)

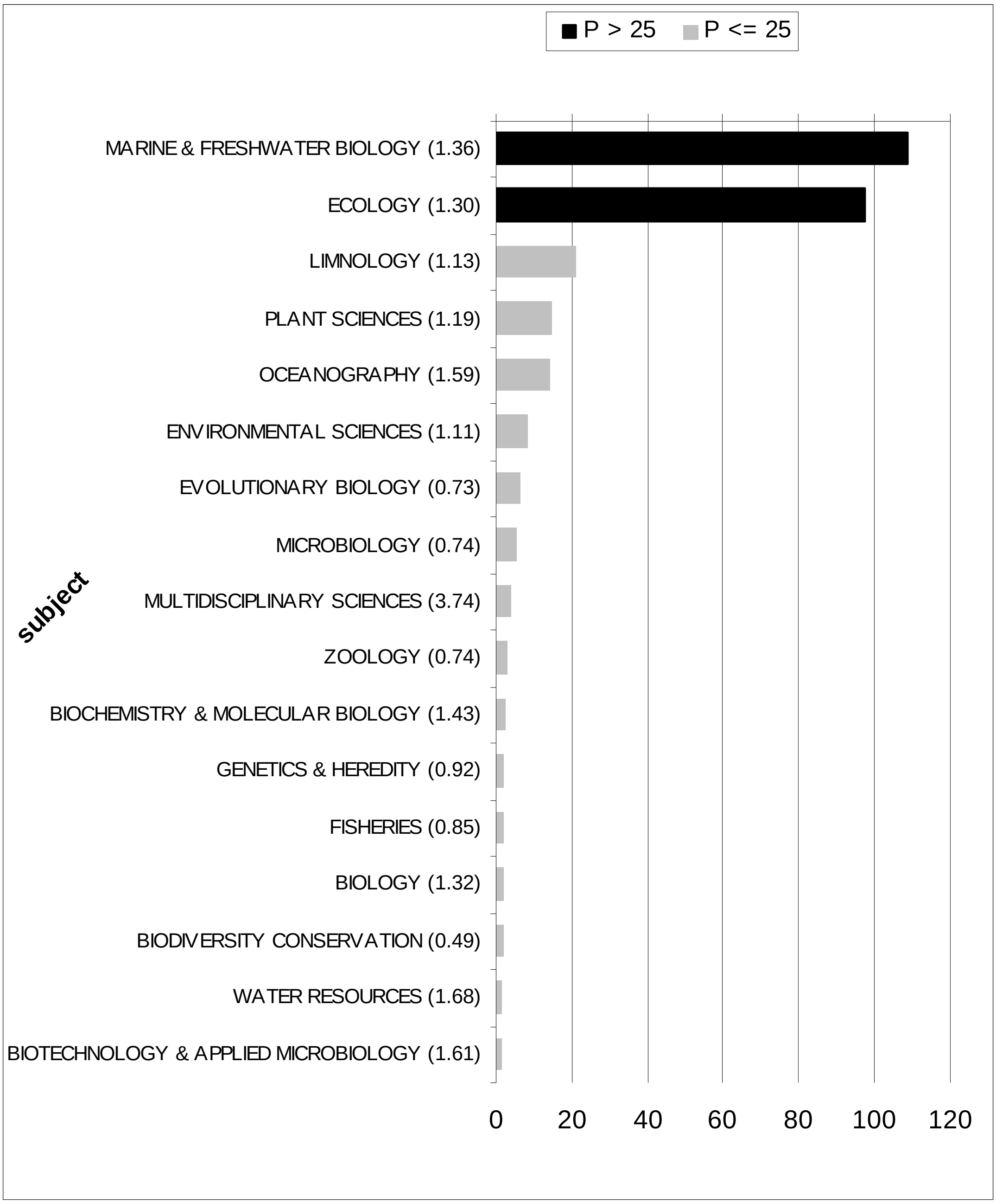
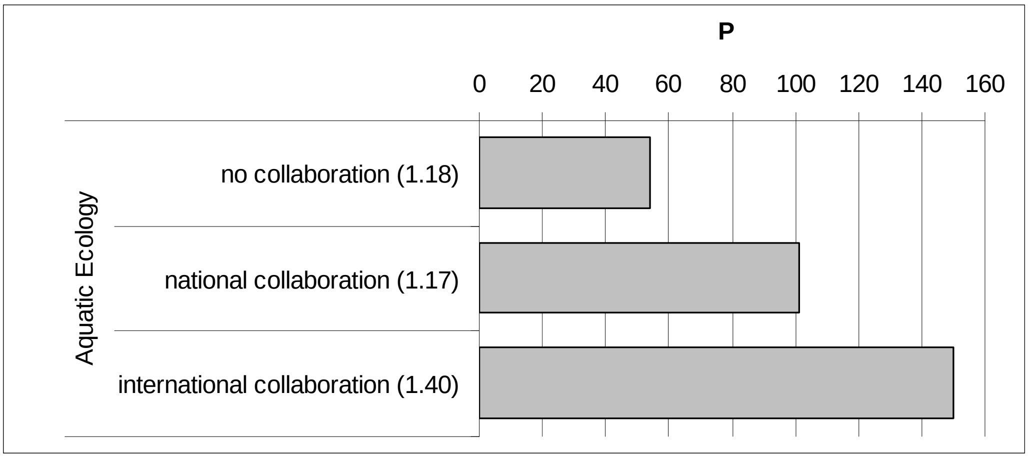


Table 5: Research profile Aquatic Ecology 2007-2010/11

subject_category	*p*	*tcs*	*mcs*	*mncs*	*pp_top_10*
ECOLOGY	49	298	6.09	0.93	0.08
MARINE & FRESHWATER BIOLOGY	31	187	6.16	1.56	0.23

Regarding collaboration, we see that Aquatic Ecology has a preference for international collaboration in which the impact is way above world average. We found out that in the most recent period (2007-2010) the impact is lowest for national collaboration, whereas for the other collaboration types, the impact is over 25% above world average.

Figure 10: Collaboration output and impact (MNCS, between parentheses) for department Aquatic Ecology (2001-2010/11)



4.2.3 Microbial Ecology

With almost 300 papers in 10 years, Microbial Ecology has a substantial output volume. The production is increased steadily of the years. The impact measured by MNCS and PPtop10% is way (around 70%) above world average. The department gets papers published in journals with a normalized impact of almost 50% above field average. In the most recent years the impact of these journals even reaches above 60% above field average. It is worth mentioning because their journals have a higher impact (MNJS) each year since 2004.

Table 6: Output and impact indicators overall and trend for department Microbial Ecology (2001-2010/11)

period	*P*	*Internal coverage*	*TCS*	*MCS*	*MNCS*	*PP top10%*	*MNJS*	*% self cits*
Microbial Ecology								
2001-2010	*281.00*	*0.88*	*2825*	*10.05*	*1.73*	*0.24*	*1.45*	*27*
2001-2004	85.00	0.86	1004	11.81	1.93	0.29	1.30	21
2002-2005	92.00	0.87	979	10.64	1.70	0.23	1.29	24
2003-2006	99.00	0.86	944	9.54	1.50	0.21	1.24	28
2004-2007	120.00	0.87	1307	10.89	1.70	0.23	1.36	26
2005-2008	126.25	0.88	1441	11.42	1.81	0.25	1.47	30
2006-2009	137.50	0.89	1434	10.43	1.75	0.26	1.58	31
2007-2010	142.00	0.90	1329	9.36	1.71	0.23	1.63	30

Figure 11: Development of normalized impact (MNCS) of Department of Microbial Ecology, 2001-2010/11 within stability intervals

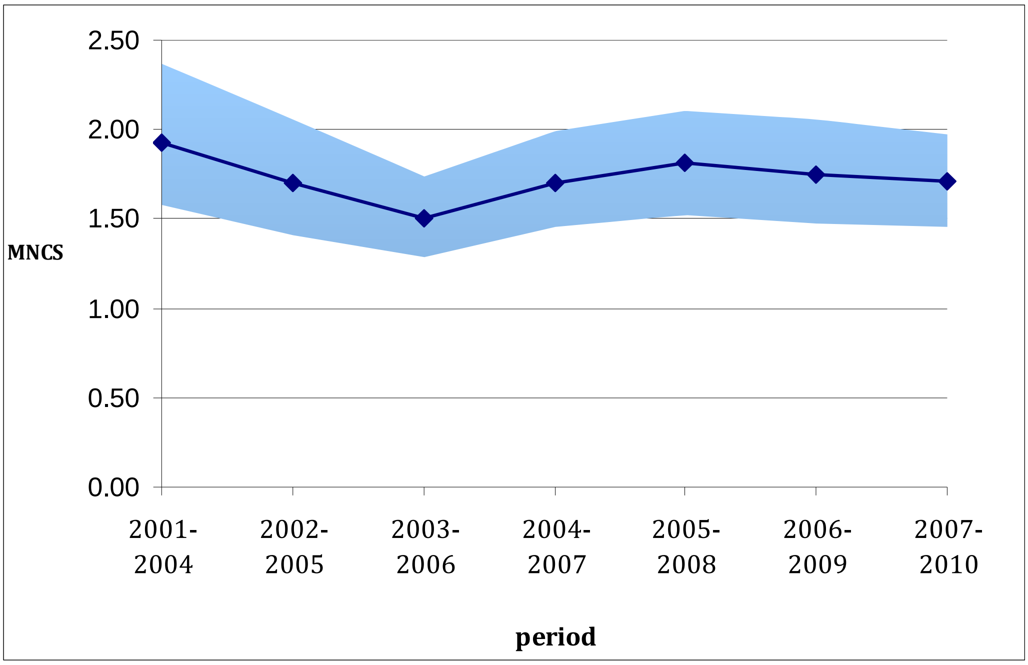
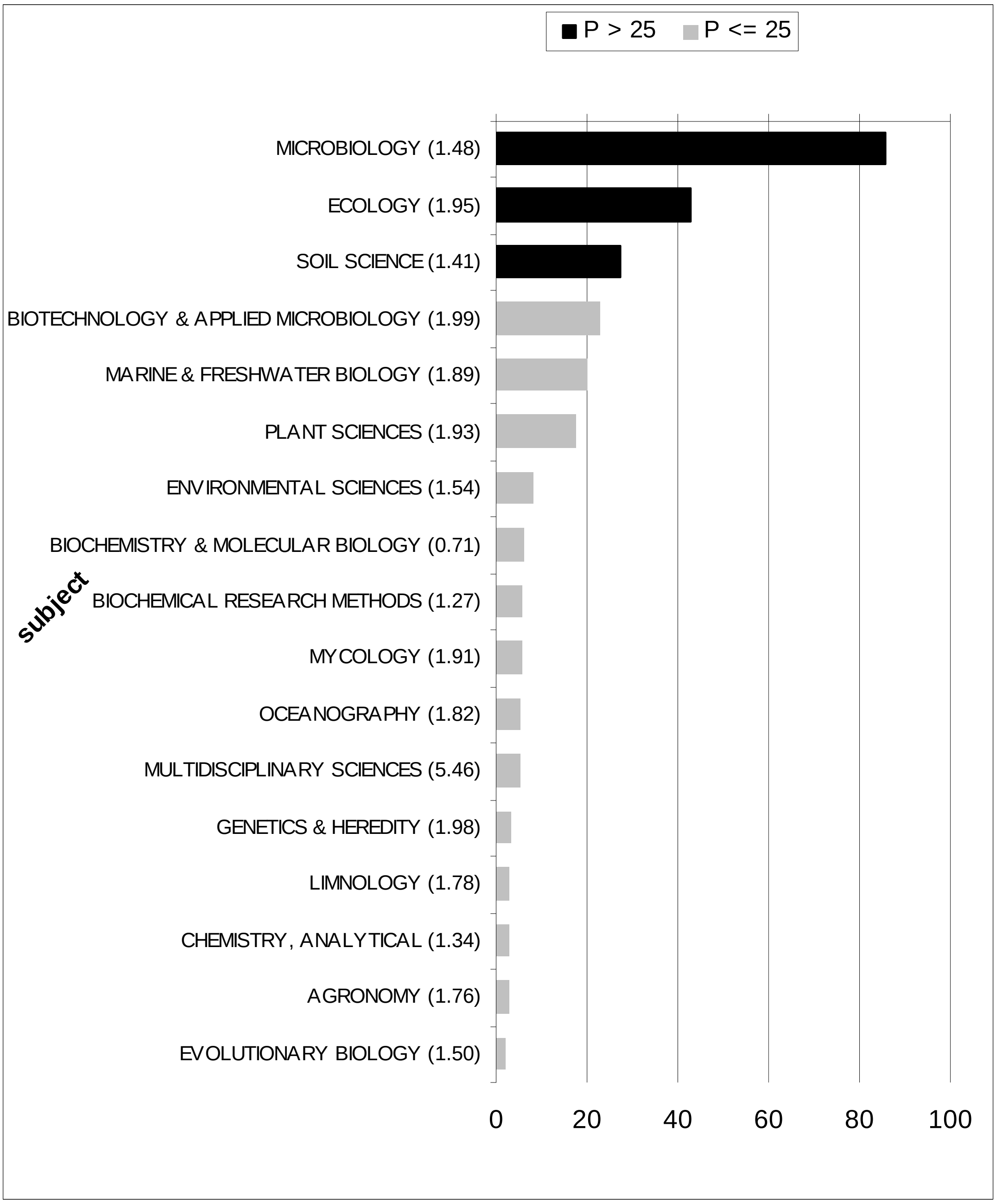


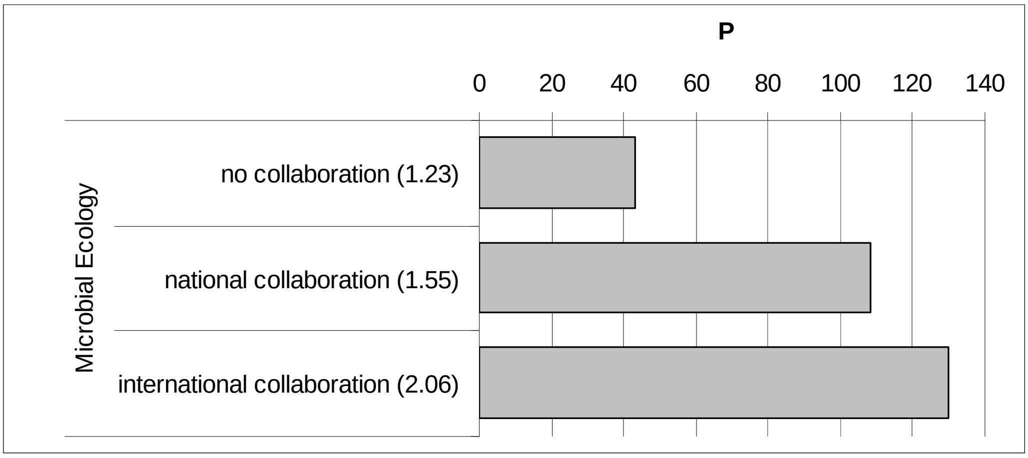
Figure 12: Research profile for Department Microbial Ecology (2001-2010/11): output and impact (MNCS, between parentheses)



The most substantial production (P>25) is in *Microbiology*, *Ecology* and *Soil science*. The highest impact is in Ecology.

NIOO-KNAW Microbial Ecology has a preference for international and national collaboration as shown in Figure 13. The impact is high in all types but most prominently for international collaborative papers.

Figure 13: Collaboration output and impact (MNCS, between parentheses) for department Microbial Ecology (2001-2010/11)



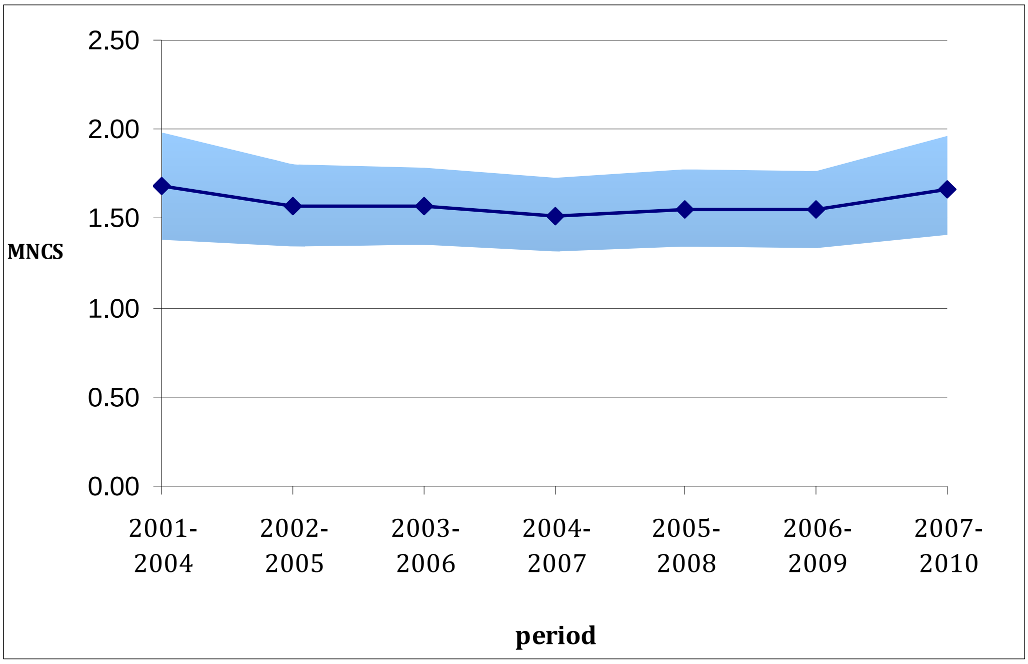
4.2.4 Terrestrial Ecology

The NIOO-KNAW department Terrestrial Ecology has the largest volume of output with almost 350 papers in 10 years. This is the only department with a stable volume over the years. There is hardly any increase of volume (P) per year since 2002. In the 10 years period the impact (both MNCS and PPtop10%) remains at a high level. Also this department publishes in the high impact journals as measured by the MNJS.

Table 7: Output and impact indicators overall and trend for department Terrestrial Ecology (2001-2010/11)

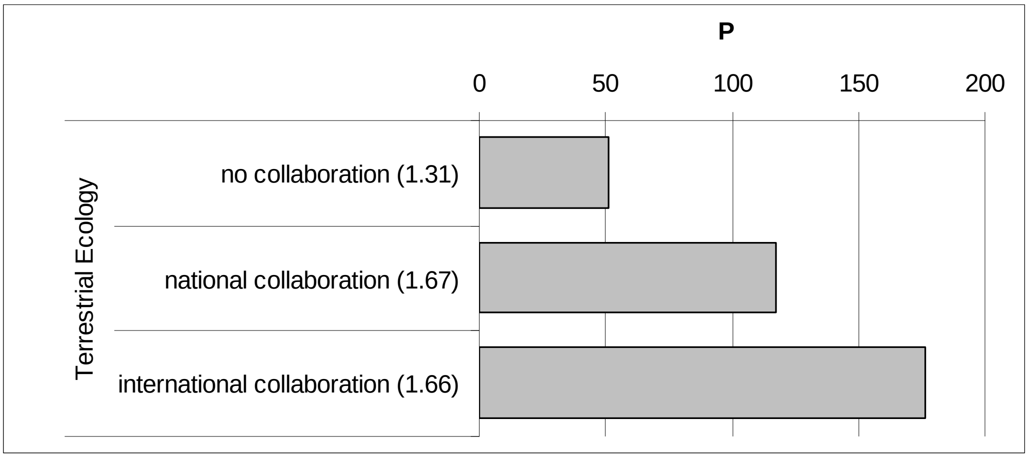
period	*P*	*Internal coverage*	*TCS*	*MCS*	*MNCS*	*PP top10%*	*MNJS*	*% self cits*
Terrestrial Ecology								
2001-2010	*345.00*	*0.84*	*2909*	*8.43*	*1.61*	*0.22*	*1.38*	*32*
2001-2004	113.25	0.80	968	8.55	1.68	0.25	1.34	33
2002-2005	142.00	0.81	1278	9.00	1.57	0.19	1.41	31
2003-2006	156.00	0.81	1387	8.89	1.57	0.20	1.39	32
2004-2007	158.25	0.82	1470	9.29	1.51	0.19	1.32	31
2005-2008	151.25	0.83	1414	9.35	1.55	0.21	1.34	32
2006-2009	145.50	0.85	1343	9.23	1.55	0.22	1.33	32
2007-2010	152.75	0.87	1293	8.47	1.66	0.22	1.42	30

Figure 14: Development of normalized impact (MNCS) of Department of Terrestrial Ecology, 2001-2010/11 within stability intervals



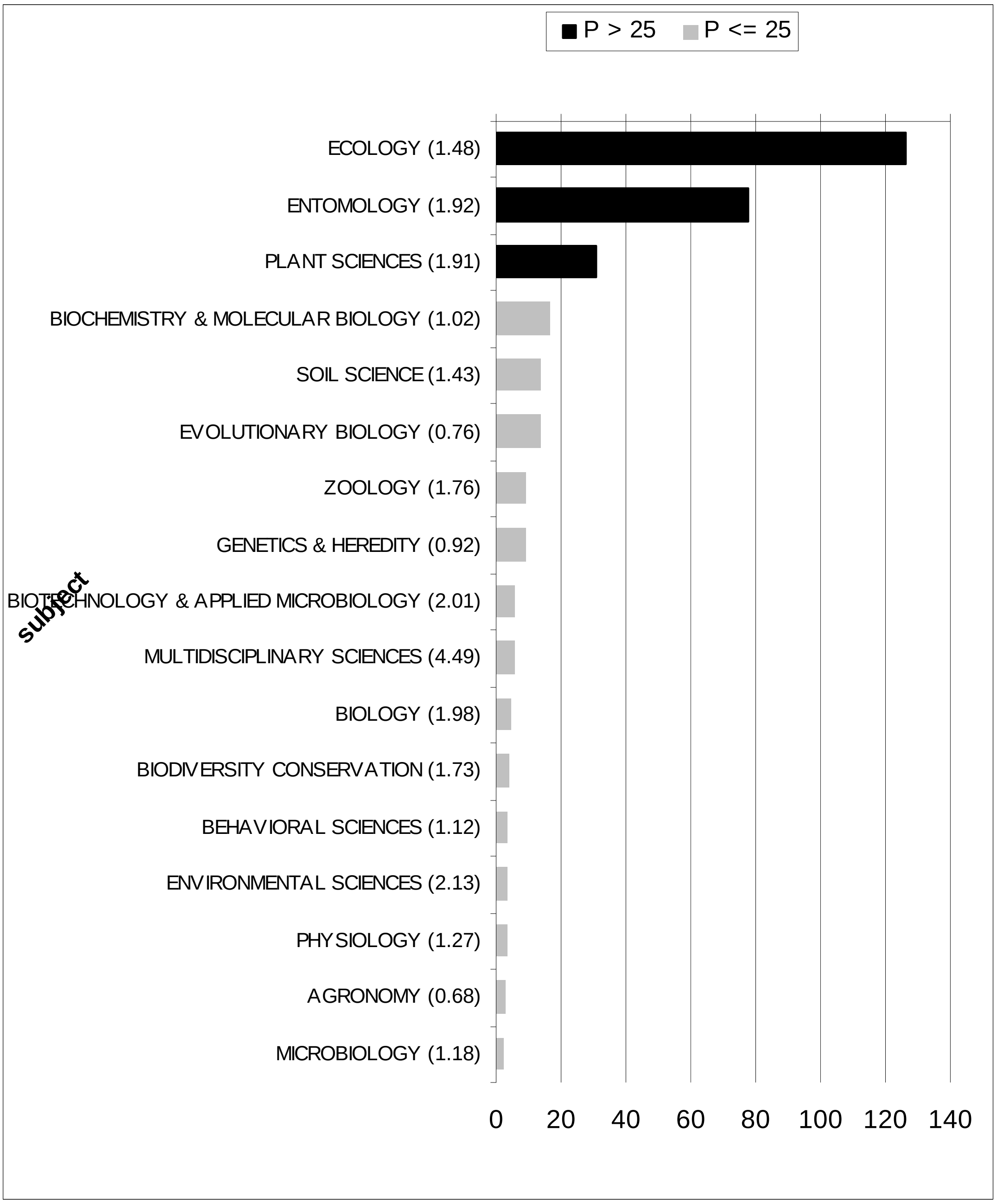
This department has a common collaboration profile with a focus on international partners. The impact is most prominently in national and international collaboration.

Figure 15: Collaboration output and impact (MNCS, between parentheses) for department Terrestrial Ecology (2001-2010/11)



The research profile of this department has a focus on *Ecology*, *Entomology* and, to a lesser extent, on *Plant Sciences*. All subject areas have an impact far above world average.

Figure 16: Research profile for Department Terrestrial Ecology (2001-2010/11): output and impact (MNCS, between parentheses)



5 Conclusions

Using our standard bibliometric indicators and analyses we evaluated the performance of NIOO-KNAW as a whole and of its departments. The performance analyses are based on the output of the institute as provided and matched against the Web of Science (WoS). We assessed that over 80% of the scholarly output of NIOO-KNAW is covered by WoS.

The entire volume of output is over 1,100 papers in ten years (2001-2010). The impact as measured both by MNCS and PPtop10% is way above world average. Moreover, NIOO-KNAW gets their papers published in high impact journals.

As for the departments, the results of our analyses show that each of them has its proper profile. All four departments have a substantial output volume over the years. Moreover, the impact in all cases is well above world average but at different levels. We dissuade comparisons of impact between departments from the following reasons. Bibliometric impact is measured by received citations and normalized against the subject fields to which the oeuvres belong. As the research profiles differ from each department to the other, the norm for each of them differs. The norm is based primarily on the subject classification of the Web of Science (WoS) which is not always appropriate (fine-grained enough) to match the specific research focus. In some cases a department's impact may be affected by this. The norm for some departments is higher than appropriate for their specific research are so that a comparison among departments would be misleading. We therefore assess the impact using the trend analysis. The benchmark for a department is its past performance rather than the performance of others. At present a more fine-grained normalization is being developed at CWTS which will be independent from the rigid and traditional subject classification and will be implemented in the near future to deal with this issue.

Regarding the results at the level of department, we found a specific character for each of them. The department Animal Ecology shows a remarkable increase of production. In 10 years the amount of papers per year is more than doubled. It has a high impact over the entire period while national collaboration plays an important part. The Aquatic Ecology department's impact seems to be affected most by the current classification in general and more in particular by a recent publication strategy shift towards general ecological journals (in the category *Ecology*). The department of Microbial Ecology is characterized by the recovery in terms of impact from 2003 onwards. Since 2005 it appears to perform at a stable high impact level. Terrestrial Ecology can be characterized as the most stable department. In the ten years of our analyses both the output volume and the impact remain almost unchanged and the latter at a high level.

References

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Garfield, E. (1979). *Citation Indexing - Its Theory and Applications in Science, Technology and Humanities*, Wiley, New York.

Glänzel, W. (1992). Publication Dynamics and Citation Impact: A Multi-Dimensional Approach to Scientometric Research Evaluation. In: P. Weingart, R. Sehringer, M. Winterhager (Eds.), *Representations of Science and Technology*. DSWO Press, Leiden 1992, 209-224. Proceedings of the International Conference on Science and Technology Indicators, Bielefeld (Germany), 10-12 June, 1990.

Martin, B.R. and J. Irvine (1983). Assessing Basic Research. Some Partial Indicators of Scientific Progress in Radio Astronomy. *Research Policy*, 12, 61-90.

Moed, H.F. (2005), *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Moed, H.F. and F. Th. Hesselink (1996). The Publication Output and Impact of Academic Chemistry Research in the Netherlands during the 1980's. *Research Policy*, 25, 819-836.

Moed, H.F., R.E. de Bruin and Th.N. van Leeuwen (1995). New Bibliometric Tools for the Assessment of National Research Performance: Database Description Overview of Indicators and First Applications. *Scientometrics*, 33, 381-425.

Narin, F. and E.S. Withlow (1990). *Measurement of Scientific Co-operation and Co-authorship in CEC-related areas of Science*, Report EUR 12900, Office for Official Publications of the European Communities, Luxembourg.

Nederhof, A.J. (1988). The validity and reliability of evaluation of scholarly performance. In: A.F.J. Van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: North-Holland/Elsevier Science Publishers, pp. 193-228.

Nederhof, A.J. & Visser, M.S. (2004). Quantitative deconstruction of citation impact indicators: Waxing field impact but waning journal impact. *Journal of Documentation*, 60, 6, 658-672.

Raan, A.F.J. van (1996). Advanced bibliometric methods as quantitative core of peer reviewbased evaluation and foresight exercises. *Scientometrics*, 36, 397-420.

Waltman, L., Van Eck, N.J., & Noyons, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629-635.

Waltman, L., N.J. van Eck, Th.N. van Leeuwen, M.S. Visser, and A.F.J. van Raan, Towards a new crown indicator: Some theoretical considerations, *Journal of Informetrics*, 2011a, 5 (1), 37-47

Waltman, L., N.J. van Eck, Th.N. van Leeuwen, M.S. Visser, and A.F.J. van Raan, Towards a new crown indicator: An empirical analysis. *Scientometrics*, 2011b, 87 (3), 467-481