The impact of science operations on science return of the Very Large Telescope

M. Sterzik[a], U. Grothkopf[a], A. Kaufer[b], B. Leibundgut[a], S. Mieske[b], F. Patat[a], M. Rejkuba[a], M. Romaniello[a], L.E. Tacconi-Garman[a]
[a]European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany
[b]European Southern Observatory, Alonso de Cordova 3107, Santiago de Chile

ABSTRACT

The operational implementation of observing programs influences the scientific return of an Observatory. More than 15 years of observations with the VLT/Paranal Observatory allow us to assess the impact of science operations and program implementation. Bibliometric parameters are used to derive program productivities and citation rates and their relation to scheduling realizations (such as service and visitor mode), program types and service mode rank classes.
In this contribution we present a set of performance indicators comparing specific program execution parameters. Results of this analysis help us to identify strengths and weaknesses of the adopted operational model, as well as possible improvements for an integrated VLT and ELT operations scheme in the next decade.
Keywords: VLT, operations, metrics, science return

1. INTRODUCTION

A decisive criterion for the success of any astronomical facility is its scientific productivity, often expressed in terms of the quantity and quality of publications it enables. Thousands of VLT programs have been executed to date, and even more papers using their data appeared in refereed journals. Since the start of operations of the VLT in 1999 the individual program properties as well as its associated bibliometric record are tracked consistently in our databases. This allows us to analyse how the scientific and operational implementation of programs affects their science return based on robust statistics.

The scientific productivity enabled with the VLT (and with other ESO facilities) in general compares very well to other observatories in terms of global bibliometrics and is widely documented [1, 2, 3]. Our analysis, however, focusses on the potentially subtle consequences of the adopted science operation model for the VLT, in particular how program modes, types, and ranks compare to each other in terms of their scientific return. The datasets utilized are coherent and statistically relevant, thus enable a comparative study that is internally fully consistent.

We have recently presented an in-depth analysis of the publication statistics coming from over 15 years of VLT observing programs [4]. Several results derive from this study, such as: (a) Large Programs are the most productive programs, even normalizing to their allocated times; (b) the publication fraction of completed programs is of the order of 60%, implying that a significant number of programs does not yield one refereed publication, while programs that result in a publication yield on average almost two refereed papers; (c) there is a trend that higher-ranked programs are more productive then lower-ranked ones, hinting that the proposal selection and ranking indeed reflects scientific merit.

In this contribution we recall the methodology, the underlying data sources, assumptions and limitations of our approach. We present additional statistical analyses to deepen the understanding of the parameters that impact the scientific return of observing programs. A correction for publication delay is discussed to enable the comparison between older and more recent data. Also papers that stem partially or entirely from archival data are considered. We derive an approximate figure of merit for observing programs by normalizing their scientific impact (as measured by citations) with the observing time required. We also start to consider the effects of sociological factors related to the research teams.

Identifying specific strengths and limitations of the underlying operational model and how it is used by the community may give us important hints to further improve the scientific return of the Paranal Observatory in the mid-term, and to get ready for an optimized implementation of an operations scheme that integrates the E-ELT and the VLT.

2. METHODOLOGY

Since the start of operations in 1999, VLT observing programs are offered in both classical (Visitor, VM) and queue (Service, SM) mode, and are implemented through different program types such as normal/open time programs, Large Programs (LP), Guaranteed Time Observations (GTO), Director's Discretionary Time (DDT), Target of Opportunity (TOO), Rapid Response Mode (RRM), Calibration, and Monitoring Programs. All program types are described in detail in the ESO Call for Proposals, which is released every 6 months for a new observing period [5]. They cover different observing modes and span the entire telescope time allocation range: GTO programs are mainly executed in VM, while Monitoring, Calibration, DDT, TOO and RRM programs are almost exclusively carried out in SM. Normal programs and LPs are being pursued in both SM and VM mode. Both VM and SM offer specific advantages. VM allows visiting astronomers to adapt their observing strategy in real-time and ensures a close relation between the Community (visiting astronomers) and the Observatory (staff). Furthermore, VM provides an opportunity for young researchers to get hands-on observing experience. SM is designed to optimally use the whole range of observing conditions according to the needs of different science cases, and enables execution of science programs requiring special conditions [4]. For both VM and SM, consistent calibration and quality control of all data gathered at the VLT allows their scientific exploitation beyond the original goals of a specific program through archive science. All these factors foster an ESO user community that is fully engaged and scientifically productive.

ESO databases contain information stemming from Phase 1 (proposal submission) and Phase 2 (observation preparation and execution). These databases exist since the start of VLT operations and most of the information is publicly available through web interfaces. Several improvements were introduced over time to make them consistent and complete, and a.o. enabled the recent analysis of operational metrics and efficiency [6]. For our purpose, the following information is used: each scheduled observing proposal with its associated program ID, Telescope, Instrument, Mode (Visitor/Service) and allocated time (nights for Visitor and hours for Service). Sometimes, programs consist of several runs, which correspond to a breakdown of the science case into identifiable subunits, and these are identified through their run IDs. Each of them may have its own allocation times, instruments, modes, etc. Observing runs are evaluated and ranked individually by the Observing Programs Committee (OPC). For SM programs, rank classes (A, B, C) are assigned by the Observing Programs Office at ESO, following the evaluations of the OPC and the VLT/VLTI Science Operations Policy. For highly ranked A programs, ESO commits all possible effort to execute the entire program, while well ranked B programs will be conducted with best effort in the requested observing period. C-rank, or filler, programs will only be executed if observing conditions do not permit to conduct other observations. VM programs are only scheduled if they are rated at the same level as SM A rank class programs.

The ESO Telescope Bibliography *(telbib)* is a database of refereed papers published by the ESO user community, which, supplemented with citation statistics from the Astrophysics Data System (ADS), allows detailed comparisons of the productivity and impact of different ESO sites, facilities and instruments [7]. In this database, the association of a paper with an ESO observing program is accomplished by linking *telbib* records to the associated ESO program IDs. Obviously, one or more ESO programs may contribute with distinct data sets to a particular paper. Cross-references between papers and programs within *telbib* is considered to be complete to about 95%, as it covers the vast majority of refereed journals in astronomy.

In order to use *telbib* to extract the bibliometric data, run IDs must be mapped to program IDs. About 2/3 of all programs considered here are "single run" programs, so the association of runs to programs is unique. The remaining programs consist of more than one run (about 20% of all programs have two runs, while in particular LPs are multi-run). In these cases, we added the time allocated to all runs and picked the rank class with the largest allocation.

For our analysis we employ the same sample of 8414 distinct programs as in [4]. They are dominated by allocations from the VLT and are scheduled between the start of UT1 operations in April 1999 (ESO Period 63) and March 2015 (end of Period 94). This study only considers observing runs scheduled at the Unit Telescopes (i.e. excludes VLTI runs, and public surveys), and only takes the program types “Normal”, “Large”, “DDT”, “GTO” and “TOO” into account. Normal programs also include so-called “Short” programs that were specifically distinguished from normal programs between Period 80 and 86. ToO programs include “Rapid Response Mode” programs, but we do not further consider “Calibration” and “Monitoring” programs in the statistical analysis due to their low numbers. The number of “mixed mode” (SM and VM runs) programs is low (<5%) and does not significantly impact the statistical analysis. In the case of mixed mode programs (combining SM and VM runs or VLT and other ESO telescope facilities), the properties of the run with the largest time allocation is assigned to the whole program.

The main program statistics and associated bibliometrics indicators are summarized in Table 1.

Table 1. Observing program statistics for the different observing program modes, SM ranks classes and types considered. A night is assumed to be equivalent to nine hours. The cumulative number of citations refers to all publications to which a given program contributed data.

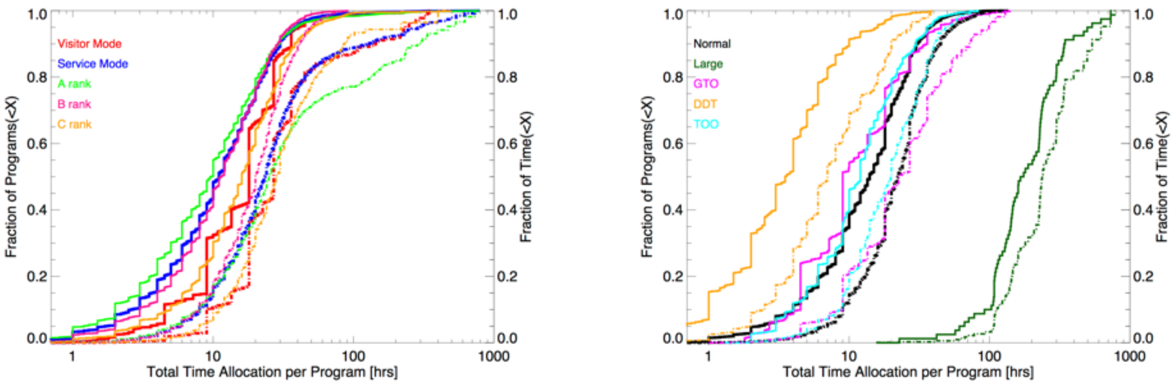
Program (Mode, Rank, Type)	Number of programs	Total Telescope Time Allocation (nights)	Number of contributions to publications	Cumulative number of citations	Number of publications per night
Total	8414	16028	11291	435018	0.70
VM	2228	5130	4211	175422	0.82
SM	6186	10898	7080	259596	0.65
A rank	2672	4807	3956	170411	0.82
B rank	2841	4515	2292	66016	0.51
C rank	673	1576	832	23169	0.53
Normal	6705	12216	7776	251368	0.64
Large	80	1862	1483	84917	0.80
GTO	498	872	960	55227	1.10
DDT	689	371	633	23533	1.70
TOO	416	672	436	19952	0.65
Calibration	17	15	3	21	0.20
Monitoring	9	20	0	0	0
Archive	*1317*		*4491*	*143339*	

Assuming that 9 hours correspond on average to one night over a year, the Total Time Allocation (TTA) for all programs sums up to approximately 4000 nights of time allocation per UT, i.e., 11.2 continuous calendar years per UT. Disregarding <5% of programs that had mixed telescopes/sites and excluding all VLTI allocations on UTs, this number is consistent with 10-15% overheads for planned technical and commissioning times and the sequential start of operations of the four UTs.

In more detail, Figure 1 displays the cumulative distributions of the number of programs (full lines), and the time allocation per program (dashed lines), for different observing modes and SM rank classes (left), and different observing program types (right). The cumulative distributions are normalized to the total number of programs, or to the TTA for all programs in each category, as given in Table 1. The median of the fractional program distribution corresponds to the median TTA in the table. The distributions for the number of programs, and their time allocations, differ. Evidently, a few programs have very high TTA, which is mostly due the (large) TTA contributions of a (few) LPs. These distributions are used to quantitatively relate telescope time (as a proxy for costs) to their science return (benefit).

The median (mean) allocation per program is 64% (27%) higher for VM than for SM programs, and the fractional distribution of time between VM/SM programs is 32%/68% (and 26%/74% in terms of number of programs). All SM programs are associated to one rank class A, B or C. A, B ranks and VM Programs have all a similar share in the total number of programs. By construction, C rank SM programs contribute by approximately 10% to the total number of all SM programs and intentionally over-schedule the queues. This procedure warrants that the available observing parameter space is filled, statistically, with suitable programs [8].
Some TTA distribution functions are skewed: in particular for A-rank programs a few Large Programs (1% in number) sum up to 13% of the total time allocated.

Figure 1. Cumulative distributions of the Telescope Time Allocation per program for the different observing program modes, SM ranks classes (left) and types (right). The different line styles refer to the normalized fractions of the number of programs (full lines), and fractions of time these programs utilized, normalized to their total allocation (dashed lines).



3. OBSERVING MODES, TYPES AND RANKS

In this section we discuss the effect of how the implementation of observing programs for the VLT in terms of observing modes, program types, and service mode ranks impacts the quantity and quality of their scientific return.

3.1 Productivity

The productivity of a program can be defined as the number of refereed publications that use its observational data, and thus describes the quantity of science return. An observing program may contribute to no publication, one publication, or more than one publication. In turn a publication often refers to a number of programs that contribute data.

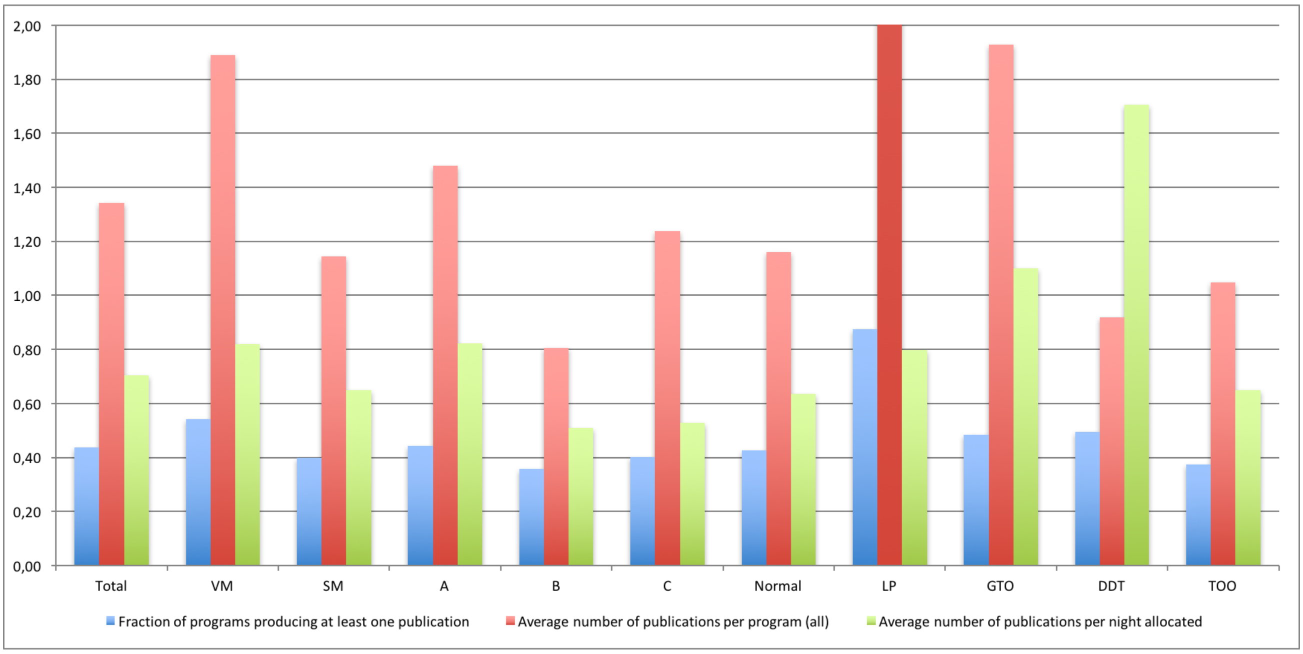
We have used the *telbib* database to correlate the ESO observing programs with those refereed papers to which they contribute data. Since the start of operations in 1999, there were 5907 distinct publications based on VLT data as of April 2015. Out of in total 8414 programs allocated, 3675 of them are mentioned 11291 times in refereed publications and contributed to them with data. During the last 16 years, observations from every allocated VLT program was used in about 1.34 publications on average (=11291/8414, the number of publications per program). On average, every VLT program contributes to 0.7 publications per night allocated (=11291/16028). At face value, a fraction of 44% (=3675/8414) of all observing programs contribute to these publications (program publication fraction).

The chart in Figure 2 shows the productivities (normalized to the number of programs respectively number of nights allocated) for different program types, modes, and rank classes. LPs exhibit the highest productivities: on average, each LP is used in 18.5 publications (note that this value is off scale). Within the same time span, program types GTO (DDT and TOO) lead to 1.93 (0.92 and 1.05) publications, respectively.
On average, VM programs result in 1.89 publications per program, while SM result in 1.14. There is a considerable spread in the productivity among A, B and C rank SM programs, which yield 1.48, 0.81, and 1.24 publications on average.

It is notable that more than half of all programs apparently do not produce any refereed publication at all. This fraction may appear surprisingly high, and it is interesting to try to understand the reasons.
A certain (but small) number of programs has never been started, although time had been allocated (e.g. TOO programs, or some SM programs), and hence the the total number of programs that received data is smaller. Secondly, typically

only 75% of the scheduled time results in useful science observations, and hence the productivities per allocated night can be corrected correspondingly [6].
Finally, a certain number of additional publications is expected to be accumulated in the future for those programs that have been scheduled in recent years only. This effect will be discussed later.

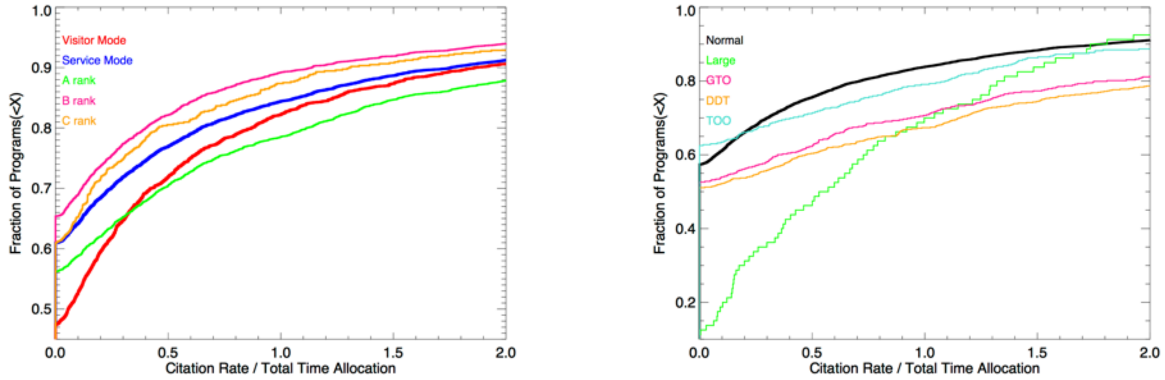
Figure 2. Productivities of various Observing Programs: Modes (VM, SM), SM ranks (A, B and C), and types (Normal, Large, GTO, DDT, TOO). The value for the number of publications of Large Programs is 18.5 (red bar, see text) and far off the chart.



3.2 Impact and Merit

The quality or scientific impact of a publication is typically measured through its number of citations. In this sense, an observing program yields impact once that program has produced at least one publication with corresponding citations. We have obtained the number of citations for a given publication from ADS, as of April 2015. All citations of all publications for a given program are summed and given in Table 1.
One may try to approximate the cost-benefit relation for observing programs by normalizing the scientific impact by the observing time required to obtain the results that have been subsequently published. Both quantities are available only approximately to us. Citation counts change over the age of a publication, and the history of citation rates for the publications is not available within *telbib*. In the following, the average citation rate for a publication is defined as its total citation count at a certain time (as of April 2015) divided by its age (defined as the difference of the time when the citation count is measured and its publication date). The assumption of a constant citation rate over the age of a publication is not correct [9]. However, this bias is introduced for all publications in the same way, and thus statistically less relevant. In case more than one publication is produced by one program, their citation rates are summed up to yield the total citation rate for a given program. In addition, we assume that the observing time required to achieve the scientific result is represented by the total time allocated for a given program (cp. Figure 1), which is an upper limit, as the actual observing time for that given program is usually smaller. A figure of merit for a given observing program may then be defined by dividing the average annual citation rate by the total time allocated for that program. The figure of merit defined in this sense is the average yearly citation rate of all contributing publications per hour of observing time allocated.

Figure 3. Cumulative distribution of the average citation rate divided by the total time allocation (in hours) for any given program. The axis of the cumulative distribution starts at larger values than zero because of the existence of that fraction of programs without publications (or zero citations).



The cumulative distribution of the figure of merit for all programs is shown in Figure 3. The figure of merit is zero for those programs that either did not produce any publication, or publications that have produced no citations. This "zero point" resembles therefore the fraction of programs that essentially have not produced any publication (thus citation). 10% of all SM and VM programs have a figure of merit above 2 (average annual citation rate per hour telescope time allocated). Notably, A rank programs have the highest figure of merit for values larger than 0.3. Within SM programs, the figure of merit for A rank programs is about 0.3-0.5 higher than for B and C. GTO and DDT programs have comparatively the highest figures of merits, with 20% of them reaching values larger than 2. Although the fraction of LPs with non-zero merit is highest, they do not reach the highest values in this metrics, because their average time allocation is comparatively high.

4. EFFECTS BEYOND PROGRAM IMPLEMENTATION

There are several additional effects that impact the science return of observing programs which are only indirectly related to the specific implementation of the science operations policy at the VLT. For example, good data quality and program completeness are prerequisites for successful data analysis and hence may impact the ability to publish. But also sociological factors like team sizes and compositions may affect how data are effectively turned into papers. Our databases allow to study the impact of some of these factors, and we present a few of them in the following.

4.1 Publication Delay

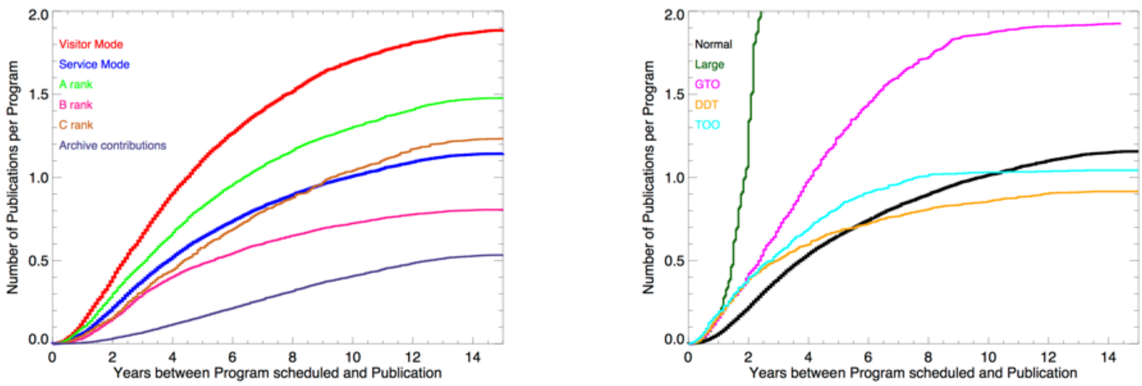
The programs and publications considered here span over more than 15 years since 1999. Obviously, turning results of an observing program into a refereed paper requires time, and more recently finished programs simply had less time to so than more mature ones. We estimate the effect of a *publication delay* in the following way. First, we relate the date of publication with the date of observations. The month of publication is available in the *telbib* database. For VM programs, the dates of observation are known from the schedule. But observations for SM programs are usually not linked to a specific date, but may span an entire observing period. Moreover, a program may last several observing periods (e.g. LP or carry-over A-ranked run). Acknowledging these uncertainties, we estimate the publication delay as the time difference between the publication time and the middle of the first scheduling period for all programs (VM and SM).

Figure 4 shows how fast the number of publications increases with time after scheduling, integrating over all programs. The highest increase of the number of publications of a program appears on average two to five years after the program has been scheduled. The smallest publication delay is present for DDT programs and 50% of all papers in this category are being published within 2 years. LPs, on average, have productivities which are one order of magnitude higher, reflected in a steeply increasing publication fraction over time. Already two years after they are scheduled they exceed two publications per program (and are off the axis limits in Figure 4). In general, the productivity for most program categories decreases significantly typically after ten years.

In principle, the growth curves allow to correct the productivities, which affect in particular “young” programs, and which are likely to yield publications in later years (at “maturity”). Assuming a simple linear growth for the first ten years (and no further growth in the following 5 years), about 2/3 of all programs are expected to contribute with an additional 50% productivity on average. This implies a *publication delay correction* of approximately 33% higher productivities, assuming the same age of all programs, i.e. a maturity after 15 years.

In Figure 4 we also consider papers that use archival data. Publications utilizing archival data boost the productivity of observing programs: about 16% of all programs considered here (=1311/8414) contribute with archival data to publications, which is about 36% of those programs that publish at all (=1311/3675). This is not surprising as a good fraction of archival papers use data of already productive programs (such as LPs or public surveys). The growth curve of archive papers is shallower than for regular programs, indicating that archival data contribute to publications *in later phases.*

Figure 4. Dependence of the integrated number of publications per program on the time difference between scheduling of the program and its use in a publication. For clarity, the curves for different program types (Normal, LP, GTO, DDT and TOO) are separated from those for different program modes (VM and SM) and SM ranks (A, B and C).

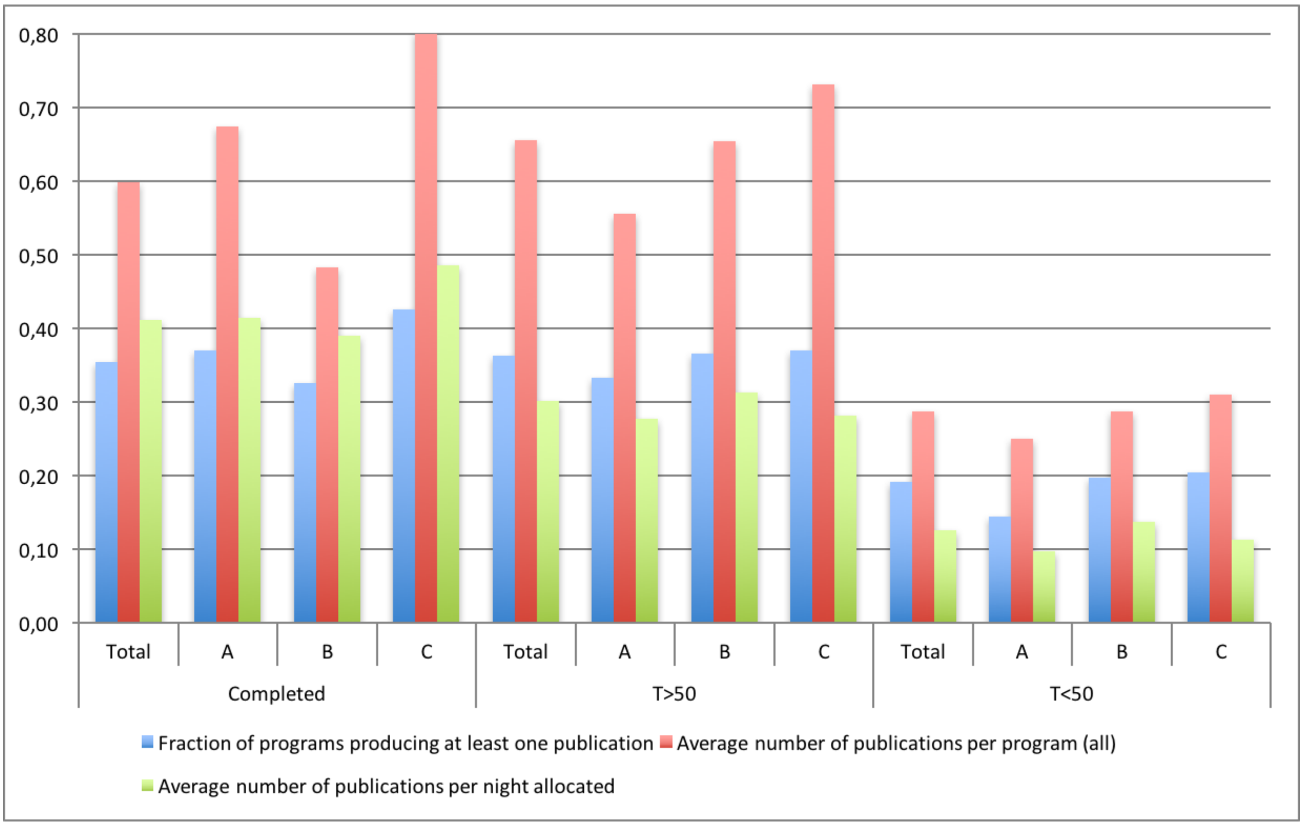


4.2 Effects of Program Completeness

The productivity of programs also depends on the completeness level of programs, in particular for B- and C-rank SM programs. Incomplete execution of observations is expected to affect the publication strategy and science impact of the programs concerned. Completeness fractions of SM programs have been systematically recorded since ESO Observing Period 78 (October 2006). An analysis of these programs has recently been presented in [6] and its impact on the productivity described in some detail in [4].

The main results are summarized in Figure 5. Normalized program productivities (as expressed by the fraction of programs with at least one refereed publication, average number of publications per program, and the average number of publications per night allocate) are displayed for the three different rank classes A, B and C, and their different program completeness levels (either fully completed, or more than 50% completed, or less than 50% completed). As the programs considered here are restricted to more recent years (and no corresponding publication delay correction was applied), the absolute numbers are smaller than in Figure 2. Figure 4 demonstrates that productivities for SM programs in all rank classes are rather similar for fully completed, and more than 50% completed programs. They are significantly lower for less than 50% completed programs. This may indicate that a completeness fraction of more than half may already allow a considerable number of programs to produce refereed papers, depending on their science case. Interestingly, B rank class programs have lower productivities as compared to A rank class programs in particular for fully completed programs. We have crosschecked that there are no statistically significant differences in the classification of the quality of the science data returned from successfully executed A- and B-rank observations. Therefore, the differences found in the performance of A- and B-rank programs may indicate that they differ in the strength of their science cases and that the program selection and the results of the ranking process does correlate with the science return of the program. C-rank (filler) programs, on the other hand, follow different science strategies and their results are turned into papers quite efficiently.

Figure 5. Dependence of productivities of Observing Programs on the completeness of SM programs for different rank classes (A, B and C). Three completeness levels are introduced: fully completed programs, programs that have been terminated with completeness levels >50%, and programs terminated with completeness levels <50%.



4.3 Effects of Team Sizes

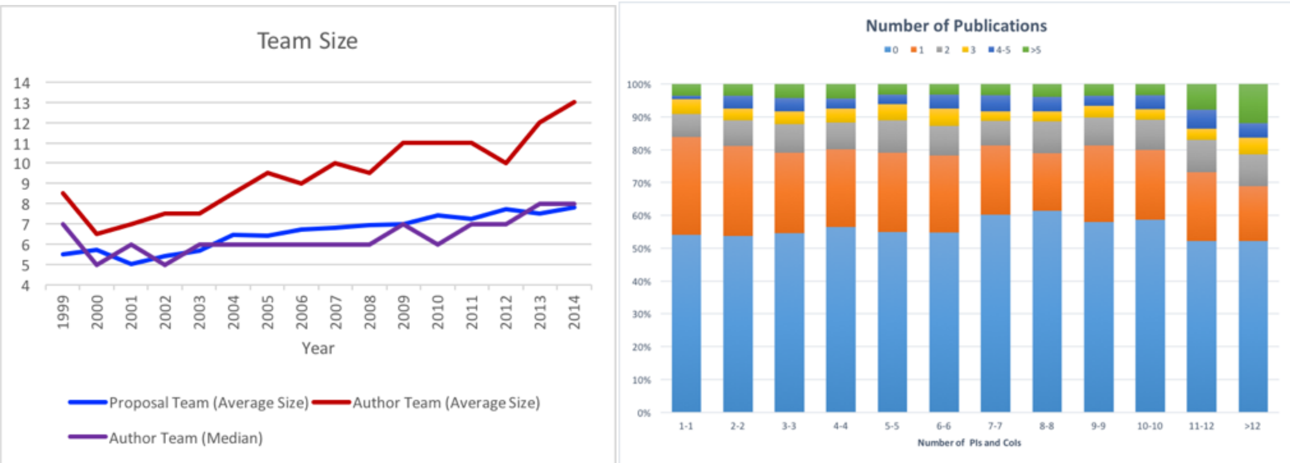
Clear trends have been observed that team sizes in science increase over the past decades, and the number of authors in publications grow also in astronomy [10]. This trend is also true for papers published based on VLT data. In the early years of VLT operations, for example, more than half of the articles were written by rather small groups of one to five authors. Over time, the fraction of teams with larger number of authors has continually increased, leading to more than 1/3 of all VLT papers in 2014 having more than 10 authors [11].

Collaborations manifest themselves already in the conception of observing programs. Our databases record the number of (Principal and Co-) Investigators per observing program, and we may therefore relate the science return of these programs (e.g. their productivity) with the number of investigators.

Figure 6 (left) demonstrates that the growth of team sizes reflects itself both in the increase of authors in scientific papers, and in the increase of investigators in the original proposals of observing programs over the last 15 years. On the right, we plot how the productivity of all observing programs (i.e. the number of papers per observing program, as defined above) relates to the number of its original investigators.

It is interesting to note that team size has only a very limited effect on the productivity of an observing program, in particular for team sizes lower than ten. Only for large teams (>12 members) a notable increase of productivity, in particular for programs that contribute to more than five papers, can be seen. These are the most productive LP and GTO teams. However, the fraction of programs that do not publish at all appears to be rather independent of the actual team size.

Figure 6. Growth of team sizes (author teams, respectively investigator teams) over the last 15 years (left). Normalized number of papers per observing program for different team sizes of the original observing programs.



4.4 Why don't we publish?

The finding that a significant number of programs does not produce any refereed papers requires further understanding. The reasons are not entirely clear from the statistical analysis alone. Although the fraction appears to be comparable to those measured at other large ground-based astronomical facilities (private communications), it is very important for ESO and its community to identify the underlying reasons. Once known, there may be opportunities to improve the service ESO is providing, and to investigate avenues that would allow to increase the scientific return of its facilities.

ESO therefore decided to poll those PIs of observing programs that did not publish, but have fully completed and highly ranked programs. The survey targets about 750 distinct PIs, with proposals approved between periods 78 and 90, regardless of their affiliation. The Survey has been launched on 6 May 2016. The selected PIs have been contacted via email by the Observing Programs Office and kindly asked to fill in a simple, web-based, multiple-choice questionnaire. On account of the need to link the replies to the proposal properties (instruments, constraints, time, science case, etc.), the responses are tagged with the corresponding Program ID, but all responses are being treated confidentially. No individual information will be published or used for any purpose other than for statistical analysis.
A very high return rate is intended in order to ensure statistically robust results. PIs have been given one month to respond, and follow-up reminders have been foreseen.
We are looking forward to making the statistical results of the survey public and discuss them in order to identify the principal causes and the most appropriate remedies.

5. RESULTS AND CONCLUSIONS

Our goal was to assess how the scientific return of VLT observing programs depends on a number of parameters that characterize the scientific and operational implementation of these programs. These parameters include the main program types (Normal, LP, DDT, GTO, TOO), the program modes (VM, SM) and rank classes (A, B, C). As proxies for scientific return we calculate program productivities, and derive an approximate figure of merit that takes into

account the average impact of a program (as measured through citation rates), weighted by its telescope time allocation. Beyond the formal program classification and implementation parameters, we consider the effects of publication delay, team sizes and program completion fractions.
Here we summarize and interpret our main findings:

1. Large Programs have by far the highest scientific productivity and impact. As expected, they fulfil their role to provide major scientific advances, breakthroughs and legacy value. But LPs also require a high investment in telescope time, which they receive after going through a rigorous scientific selection. Normalised by the allocation time, their impact (citation rates) are still competitive with all other program types. Thus LPs are proven to be a highly valuable asset in the strategic distribution and implementation of VLT programs.
2. Most of the telescope time available at the VLT is allocated for the execution of Normal Programs. After correcting the publication delay, their data are used in on average 1.5 refereed publications after 15 years, with mean telescope time allocations of 2 nights per program.
3. DDT programs perform in particular well for smaller time allocations: with a small investment of telescope time, they answer a specific, "hot", scientific question that can be published immediately; while their absolute impact (in terms of citations) remains limited, their merit (in terms of citation rates per time allocation unit) is high.
4. GTO programs have on average a higher impact than normal programs, supporting their role as pathfinder using novel instrumentation for cutting-edge science cases. Also their merit is comparatively high.
5. Also TOO programs have on average a higher impact than normal programs. And they are also productive, considering that TOO programs are in general allocated to respond to unforeseeable events, and several of them have not been triggered in a given observing period.
6. VM programs exhibit high productivities and impact, in particular for normal programs with telescope time allocation of a few nights. In this parameter range, the specific strengths of VM allocations pay off: the visiting astronomer can optimize the observing strategy and implement back-up programmes to adjust in quasi real-time to changing observing conditions.
7. Within SM programs, A-, B- and C-rank programs exhibit differences in their productivities, and impact. Upon "maturity" (i.e. considering the publication delay correction) A-rank programs yield, on average, 2 refereed publications per program. C-rank programs produce typically 1.6 publications per program, while B-rank programs reach 1.1, which is 45% lower than for A-ranks.
8. By design, A-rank programs are more complete than B- and C-rank programmes. As expected, the number of publications per program, and per telescope time allocated, increases with completeness fraction for all ranks. Statistically, most B-rank programs lose their programme execution competition to A-rank programs in the same, demanding, observing constraint conditions. This is one likely reason for the relative underperformance of some B-rank programs. But even considering only a sub-set of fully completed, normal, programmes, the productivity of B-rank programs in terms of number of publications per program is 37% lower than for A-ranks. Proposal selection and ranking appears indeed indicative of their scientific merit.
9. C-rank programs perform well in terms of normalised productivity and impact. They are fully worth the "effort", respectively the investment in telescope time. By designing them as filler programs, C-rank programs do not compete for any observing constraint parameters.
10. Even some incomplete programs utilize the obtained observations in publications. The impact of program completeness on its scientific productivity likely depends on the individual science cases.
11. The publication ratio for all VLT programs (i.e. the number of programs that publish at least one refereed paper, divided by the number of all VLT programs scheduled) reaches 58% after 15 years. Depending on the mode and type of a program, about 33%-50% of all scheduled programs *do not* contribute to refereed publications. This fraction may appear surprising and high, and we hope to understand the reasons better with the feedback from a dedicated user survey.
12. The contribution of archival data to publications is steadily increasing. More than one third of all programs that are used for publications also contribute to publications with archival data. It can be expected that an increased science value of the archive, e.g. through the massive availability of science grade data products, further increases the publication fraction. One may also expect that in case science grade data products are available close in time with the observations, the publication delay of the results may decrease, and thus help to increase the publication pressure for the PIs of the original program.

Overall, the ESO science operation policy and implementation appears to be fully adequate to enable a healthy science return of the facilities. Program types, modes and ranks open and adequately cover the available parameter spaces for their respective domains.
Attention must be given to the implementation of B-rank programs, and in particular to facilitate an increase of their completion rates for some programs. C-rank programs exhibit a relatively strong performance. In the light of making optimal use of the available observing time, it appears reasonable to strengthen this channel in SM.
Together with the community we try to better understand the reasons why a significant fraction of VLT programs does not lead to results yet published in refereed journals.
The statistical analysis presented here and in [4] and [6] are only one further step towards a better global understanding of the complexity of the various processes that, at the end, lead to an increase of scientific knowledge. The strength of individual science cases together with the capabilities of the facility, but also sociological factors within the science teams may very much influence the scientific return, stretching well beyond the scope of the present analysis.

REFERENCES

[1] http://www.eso.org/sci/libraries/edocs/ESO/ESOstats.pdf
[2] Trimble, V., Ceja, J.A., "Productivity and impact of astronomical facilities: Three years of publications and citation rates", Astr. Nach., Vol.329, Issue 6, 632 (2008).
[3] Crabtree, D., "A bibliometric analysis of observatory publications for the period 2008-2012 ", *Proc. SPIE* 9149, Observatory Operations: Strategies, Processes, and Systems V, 91490A (July 28, 2014); doi:10.1117/12.2054058; http://dx.doi.org/10.1117/12.2054058
[4] Sterzik, M., et al., "The Scientific Return of VLT Programmes", ESO Messenger 162, 2 (2015).
[5] ESO Call for Proposals, e.g. the most recent one at http://www.eso.org/sci/observing/phase1/p98/CfP98.pdf
[6] Primas, F., et al., "Fifteen Years of Service Mode Operations: Closing the Loop with the Community", ESO Messenger 158, 8 (2014).
[7] see also http://telbib.eso.org/
[8] Silva, D., "Service mode scheduling: a primer for users", ESO Messenger, 105, 18 (2001).
[9] Crabtree, D. and Bryson, E., "The Effectiveness of the Canada-France-Hawaii Telescope", JRASC, Vol. 95, 259 (2001).
[10] Milojevic, S., "Principles of scientific research team formation and evolution", PNAS, vol. 111, no. 11, 3984-3989, (2014).
[11] ESO Annual Report 2014, page 27.