A Model for Budget Allocation in Multi-Unit Libraries

Anish Arora (aarora1@uiuc.edu)
Diego Klabjan (klabjan@uiuc.edu)
Department of Mechanical and Industrial Engineering
University of Illinois at Urbana-Champaign
1204 W. Green St., Urbana, IL 61801
Tel: (217) 265 8218 Fax: (217) 244 6534

Abstract

This article addresses the problem of allocating funds for acquisition of periodicals among several interrelated units of an academic library. The paper outlines an optimization model for the problem with an objective of maximizing the usage of periodicals over all library units subject to a single collections budget. It further gives an insight into a decision support system in MS Excel based on this model, and it illustrates its successful application to the Library of the University of Illinois at Urbana-Champaign.

1 Introduction

Budget allocation is a core problem faced by both public and private academic libraries all over the world. The cost of library materials has mushroomed during the past decade while increases to the Libraries' collections budget have not kept pace. Kean [8] reports a 10% per year average increase in price for all U.S. periodicals from 1990 to 2001. In addition, the same figures for periodicals in the sciences and medicine averaged 13.5% from 1995-2001. On the contrary, the average yearly increase to the collections budget has not increased significantly in most libraries. The dramatic difference between the rate of increase for periodical prices and the Library's collections budget illustrates the severe magnitude of the library budget allocation problem. In addition to double-digit percentages for periodical price increases, Libraries face other challenges associated with library acquisitions. The Academia is currently witnessing an exponential growth in the amount of scholarly research published. Therefore, the issue is not one of supporting only the increased cost of what the libraries already have, but also of being positioned to acquire newly published works representative of scholarly research. Complicating the effect of these trends is the growth of electronic resources. Predictions that electronic publication will usurp print have not yet become fact, however, as this area of publication grows, the demand on the budget will increase.

Of all of these challenges, the periodical price increase problem is the most severe, because it has wreaked havoc with the Library's collections budget. To address this problem, libraries all over the world are considering periodical cancelations, which has left the libraries with a daunting task of deciding which periodicals to cancel. The University of Illinois at Urbana-Champaign (UIUC) Library has not been excluded from this growing budget crisis. In 1966, UIUC's Library collections budget was the highest of any university in North America. With the price of scientific periodicals escalating, combined with the effect of inflation and political decisions, by 1990s, the budget increases were well below the rate of inflation. In fiscal year 1997, the UIUC's Library collections budget had slipped to 16th in the U.S. and in 1998 the library faced its worst budget crisis ever with a shortfall of nearly $800,000 (Report of the Library Allocation Steering Committee [14]).

Under one typical model, a large academic library is composed of several library units, e.g. Engineering, Mathematics, Business, etc., which are interrelated and have a single collections budget. A key problem faced by such a library in the context of insufficiency of funds is to provide a fair basis of allocating these funds to the library units. Most of the work done in this field so far are *formula* based approaches, Budd [3], which are fairly popular by virtue of their comprehensibility and ease of altering and adapting for different situations. German and Schmidt [5] give specifics on the current budget allocation practice at the UIUC Library and they also address the need for a new model as a replacement for the existing one. The UIUC Library currently uses a formula based budget allocation model, which allocates 60% of new monies proportionally on the basis of 4 factors, viz. the number of full time equivalent faculty, the number of undergraduate and graduate instructional units taught, the number of Masters degrees awarded and the number of Ph.D. degrees awarded. It took several years of discussions and debates to reach an agreement on what factors to include and how to weigh them in the formula, which has nonetheless been the focal point of contention and has drawn criticism from various parties. Currently the model weighs all the factors equally. The remaining 40% of the new monies are allocated based on professional judgement, which is captured from scores gathered from a questionnaire filled by librarians on the Budget Subcommittee. There was seldom a consensus among scores assigned by the members of the subcommittee. Although the current approach of allocating the monies is an improvement over the past practice, however whether the model is instrumental in generating an allocation which reflects current campus priorities and how responsive it is to the changing disciplinary needs is questionable. One of the recommendations which have been proposed by committees associated with the Library, was to reduce dependence on the formula based approach and to replace it by alternative models, which would provide a fairer basis of allocation.

The authors have developed and implemented a model that facilitates the allocation of the periodical collections budget to the library units. The main objective of this model is to make an allocation that maximizes the *usage of periodicals*. A possible measure of usage is the number of publications and citations in periodicals by faculty. The model uses forecasting to measure the usage based on historical data and studies time-trends of various factors, which accrue towards the usage. By slightly modifying a family of constraints, the model can be used by librarians to facilitate acquisitions decision making. A similar quantitative model can be developed for acquisition of monographs.

By far the most popular approaches to library budget allocation are formula based allocations, E. Bourgeois and Natesan [4], Lowry [10], Budd [3], M. Niemeyer and Slattery [12]. In these approaches the allocation is determined proportionally based on several factors such as the number of instructional units, degrees granted, the number of full time employees, etc. However it is hard to determine which factors to include in the formula and how to weigh them with a rationale which justifies these decisions. There have been attempts of some scientific and operations research based approaches as well. Zilla [16] addresses a related problem of allocating a single budget to university units. She uses simulation to estimate the future needs of the departments and she develops a minimum cost network flow model for budget allocation. Wise and Perushek [15] use goal programming to determine the library allocations. They identify a set of goals, e.g. not exceeding the budget, allocating based on the number of faculty, etc. In addition they assign a priority and a weight to each goal. Another approach is discussed in Gleeson and Ottensmann [6], which describes a decision support system for acquisition of new materials implemented at the Monroe County Public Library in Indiana. However, this being a single unit library and holding primarily monographs, makes it significantly different from the problem under consideration. In addition an allocation model for a public library is unlikely to fit the needs of an academic library. Their model is usage based and they use forecasting to estimate the usage of monographs in the time horizon. Promis [13] illustrates the budget allocation strategy implemented at the University of Arizona Library. Their allocation is based entirely on reports obtained from the university databases on departmental information (reflecting potential users), usage information (such as circulation statistics, interlibrary loan statistics, library holdings, etc.) and some external data sources such as the worldwide book and serial cost information.

Section 2 gives a description of the model. This section describes the input data and how the data is used to obtain the usage of periodicals. Section 3 deals with the model implementation and the decision support system. Computational results are given in this section.

2 The Allocation Model

This section illustrates a new model to solve the problem of allocating funds to the various library units. It is assumed that the library has a single budget for periodicals that has to be allocated to several units of a multi-unit library. The objective of the model is to allocate funds to the units in such a way that the usage of periodicals across all the library units is maximized. One of the challenges apparent here is how to quantify usage. Data obtained from reshelving seems to be the best option, since it gives a firsthand information on the usage. However standardizing reshelving procedures in all library units and implementing it successfully is not an easy task. Moreover, implementing reshelving on a large scale is a costly operation as not only will it necessitate the use of expensive information systems and huge database systems but also require the library to employ a considerably large staff to implement and maintain these systems pervasively. Consequently, alternative measures of usage are taken into consideration. After consulting librarians and other staff members of the library, it was inferred that the number of publications in a particular journal by the faculty and the number of times it has been cited by the faculty is a reasonable measure of the journal's usage. Furthermore, considering the fact that this information is readily available, it seems to be the most viable option for quantifying the usage.

The data used by this model is obtained from the ISI database by ISI Thompson Scientific [7]. This commercial database contains information on journals and reports the number of citations and publications in each journal by the faculty of an academic institution since 1981. Since the funds have to be allocated for future use, it is much more meaningful to forecast these numbers for the forthcoming fiscal years and use them as the basis of usage. Details on computing the usage are given in Section 2.1.1.

The ISI database categorizes the journals based on subject areas or categories such as Mathematics, Organic Chemistry, Psychology, etc. The model uses forecasted usage values as the objective function and it makes decisions on how many journals to purchase in each category. The constraints include a given budget figure, and upper and/or lower bounds (if any) on the amount of funds which can be allocated to a library unit.

The basic steps in building the model are enlisted below. Each of these steps is discussed in detail in the section that follows.

- Quantifying the usage of periodicals.
- Creating subcategories. Since usage in a category may vary substantially from a periodical to a periodical, each category is split into two subcategories, each one of them having small variance in usage.
- Forecasting usage in each subcategory.
- Formulating the model.
- Computing the allocation and analysis of results. The allocation to each library unit is calculated based upon the number of periodicals in each category selected by the model, the average cost of the periodicals in each category and based upon the categories that are held at the library unit.

2.1 Methodology

2.1.1 Quantifying Usage of Periodicals

As mentioned earlier, the usage of periodicals is based on the following information, which is obtained from the ISI database.

- The number of citations by the faculty and
- the number of publications by the faculty.

Discussions with librarians have revealed that this data would be a good measure of usage. Together with the librarians, the authors performed a preliminary analysis to check if the data are actually indicative of usage. They observed that for subject areas that are historically strong at UIUC, there are clearly a significant number of journals having an above average number of citations and publications by the UIUC faculty. On the other hand, subject areas that are relatively less developed had relatively low figures. Although the results of this analysis were satisfactory enough to justify that there is clearly a correlation between usage in a subject area and the number of citations and publications in that area, the exact nature of this relationship was unknown.

The second step was to combine the data for publications and citations as a measure of usage. A general strategy is to assign weights to the number of citations and the number of publications and use the combined data in the model. Each subject area can have its own weight. This is somewhat similar to what is done in formula based approaches. However, due to the large number of subject areas, assigning weights to each of them is extremely difficult. At the same time, the librarians tend to be more inclined towards the number of publications as a more reliable measure of usage as compared to the number of citations. Approximately one third of the journals have zero publications and therefore they would be treated equally if only publications were considered. For these journals it is desirable to include the number of citations as a measure of their usage. Given a category, for every journal in the category, the total number of citations and publications is calculated over all years. Next, the average ratio f of the total number of publications to the total number of citations over all the journals in the category that have at least one citation and publication is computed. Thus f estimates that for every citation there are f publications in the category. For each journal in the category with zero publications, the *usage value* for every year is calculated as the number of citations in the year multiplied by f. In this way the authors combined publications and citations and obtained the usage value for almost every journal. For the journals that have at least one publication in all of the years, they define the usage value as the number of publications in a year. Only about a dozen journals have no data on citations and publications.

Note that the presented procedure can easily be extended to incorporate weights for citations and publications.

2.1.2 Creating Subcategories

It is observed that periodicals in many categories have a remarkably large variance of *cumulative usage values*. The cumulative usage value of a periodical refers to the sum of usage values over all the years for which the data is available. This is expected due to the fact that some periodicals are more popular than others, represent a greater audience and contribute to academic research more significantly than others and hence have relatively higher cumulative usage values than other periodicals in the category. In case all periodicals in a category are treated as equal and it is assumed the average cumulative usage value among all of the periodicals is representative of the cumulative usage value of every periodical in the category, there is a risk of eliminating some periodicals from the budget plan which may have relatively high usage and therefore may be vital for the library's collection. In order to overcome this problem the authors split each category into two subcategories, representing periodicals with relatively high and relatively low cumulative usage values respectively. In other words, the variance of cumulative usage values in a category may be extremely high but the two subcategories have a much lower variance. An extensive analysis of the usage data brought to light some interesting observations.

• Most of the periodicals in each category have almost the same cumulative usage value, but there are a handful of periodicals with significantly higher cumulative usage values, i.e. *outliers*.

• The average cumulative usage value in a category represents neither periodicals with low cumulative usage values nor those with high cumulative usage values.

• The proportion of periodicals with relatively high cumulative usage values varied from category to category and therefore splitting categories in predefined proportions or based on a predefined cumulative usage value would not be a good strategy for all categories. It is thus necessary to design an algorithm for creating subcategories, which can be used pervasively.

Based on these observations, the authors designed an algorithm for splitting the categories, which is illustrated in Figure 1. The algorithm splits each category into two subcategories. Initially all the periodicals in a category are assigned to subcategory 1 which represents journals with low usage. Subcategory 2 corresponds to journals with high usage and is empty initially. In every iteration the mean and the standard deviation of the cumulative usage values of all periodicals in subcategory 1 is calculated and it is checked if the standard deviation is less than the mean. If this condition does not hold, the periodical with the highest cumulative usage value in subcategory 1 is moved to subcategory 2. This procedure is repeated until the condition holds true.

Computational experiments have shown that the two created subcategories have remarkably distinct average cumulative usage values, and that the average values approximate well the cumulative usage values of all the periodicals in the subcategory, i.e. within each subcategory the variance in cumulative usage values is within acceptable limits. Note that the periodicals in subcategory 1 represent the true decision making variables since all the periodicals in subcategory 2 have high cumulative usage values and therefore are extremely likely to be acquired. Splitting each category into more than 2 subcategories would create an even smaller variance, however it would increase the forecasting error since the cumulative usage values in each subcategory would be low.

2.1.3 Forecasting Usage in each Subcategory

Given that forecasting for a larger sample size is more reliable and accurate than forecasting of individual items (journals in this case), the authors forecast usage for each subcategory.

Let t be a positive integer representing the index of a year such that $t = 0$ represents the foremost year (1981 in our case), and the successive years are represented by consecutive values of *t*. Consider a subcategory i and let A_{ti} be the sum of usage values of all periodicals in subcategory i in year with index *t*. Let F_{Ti} be the forecasted usage in subcategory i in the future time period indexed by *T* for which the allocation is sought. The authors use the *exponentially smoothed time-series model* , Makridakis and Wheelwright [11], to compute the forecasted usage. The forecasted value is computed as

$$F_{Ti} = \alpha_i \sum_{t=1}^{T-1} (1-\alpha_i)^t A_{(T-t)i} + (1-\alpha_i)^T A_{0i} \qquad 0 \leq \alpha_i \leq 1, T \geq 1,$$

where α_i represents the best damping factor for subcategory i. For every α_i = 0.1,0.2,...,0.8,0.9, the lowest mean square error is calculated. The damping factor that gives us the lowest mean square error is chosen.

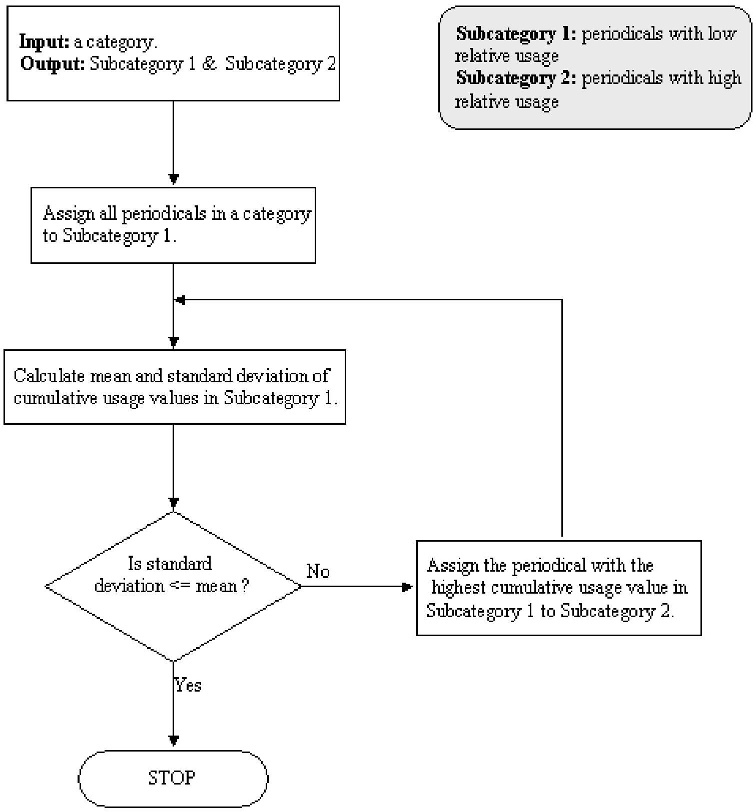

Figure 1: Algorithm for creating subcategories

Due to the relatively large number of subcategories, it is difficult to use other forecasting methods such as the Box Jenkin's methodology.

2.1.4 Formulating the Model

Ideally it would be desired to be able to make decisions on which specific journals to buy. Such an approach has two drawbacks. First, forecasting would be difficult and likely to yield large errors. Second, it would be extremely arduous in such low level decisions to capture all the judgmental reasoning that the librarians use. Keeping these issues in mind, the authors comprehended that there was a need to group the journals. Since the ISI database has a predefined set of categories, they use these categories in the model. Since each category is split into two subcategories, they actually use subcategories as the basic unit in the model for which decisions are made, i.e. the model is not capable of deciding *exactly which* journals must be purchased, but it decides *how many* journals must be purchased in each subcategory. However, since the ISI classification assigns many journals to more than one category, it necessitates a different approach to aggregation to avoid double counting journals.

A journal is *cross-listed* if it is assigned to two or more subcategories. The journals that are not cross-listed correspond to the journals that are assigned to exactly one subcategory. It was observed that many journals are assigned to either a single subcategory or to exactly two subcategories, and that approximately

7% of the journals are assigned to more than two subcategories. The purchasing decisions of the journals that are not cross-listed do not affect the number of journals to be purchased in any other subcategory. Around 40% of the journals are assigned to exactly two subcategories, i.e. they are cross-listed with exactly two subcategories. These journals affect the purchasing decision in exactly one other subcategory. The remaining journals are cross-listed more than twice. They represent journals such that their purchasing affects purchasing decisions in at least three subcategories. Journals in this group are treated individually, i.e. they are represented by binary decisions. For these journals as well, the authors use the forecasts for subcategories.

The detailed model is illustrated in the appendix. It embeds three types of decisions.

1. The number of journals to purchase in each subcategory from among all journals that are not cross-listed.

2. For every pair of subcategories, the number of journals to purchase from the set of all those journals that are cross-listed in the pair.

3. To purchase or not to purchase a journal that is cross-listed more than twice.

The objective function maximizes the usage forecasted for the forthcoming fiscal year, as explained in Section 2.1.3. The model also allows weights that prioritize categories. Note that these weights are different from those discussed in Section 2.1.1. In addition the model has the following constraints, which ensure that the model generates a legal allocation.

- A constraint to assure that the model does not allocate more funds than the available budget. In this constraint we use the journal prices, since for a given category, the average journal price multiplied by the proposed number of journals for purchasing yields the expenditure for that category.

- Constraints to guarantee that no more journals than the maximum available in each subcategory are allocated.

- Constraints which prevent the model from allocating more cross-listed journals than those available.

- Constraints to ensure that the model does not allocate to a library unit more funds than a given lower and upper bound.

2.1.5 Computing the Allocation and the Analysis of Results

This is the final step in the budget calculation process for periodicals. The model yields the number of periodicals that should be purchased in each subcategory however the goal is to allocate funds to each library unit. The monies allocated to each library unit are computed based on constraints (5), which are given in the appendix. The decision support system, which is described in the next section also generates various statistics and charts such as the percentage of allocated periodicals and funds, in order to facilitate the decision making process. Librarians are interested in two additional results. They want to know how many journals should be ideally purchased in each category. They are also interested in knowing the number of journals that should be purchased in each category regardless of journal prices (i.e. taking into account only the usage). These questions can be easily answered by making a few changes in the model which are discussed in the appendix.

2.2 Regression Based Forecasting

An alternative approach to time-series forecasting is regression based forecasting, which is described here. The assumption is that the usage is a function of several factors and can be represented by a linear equation, where these factors are independent variables and usage is the dependent variable. Some of these factors for a department are: the number of full-time employees, the number of research assistants, the amount of grants and contracts per faculty member, the number of graduate students, the number of instructional units, etc. UIUC has the historical data for all these factors and for every department. It is then possible to construct a multivariate linear regression model based on minimizing the autocorrelation and heteroscedasticity among the independent variables.

After obtaining the regression equations, one can forecast the usage for each department using values of these factors for the forthcoming fiscal year, in the respective regression equation. This approach would more accurately model sudden changes in the departments, such as new hiring policies, increase or decrease in funds available for research, etc. For example, if it is known that a department will significantly increase the number of full-time employees, the regression model can effectively capture this. Since the time-series forecasting model was sufficiently accurate, a regression based forecasting system was not implemented.

3 The Decision Support System in MS Excel

This section gives an insight into the decision support system (DSS) developed by the authors. The system is developed within MS Excel by using a a set of macros written in VBA, Albright [1]. The user interacts with the DSS via a graphic user interface (pull-down menus and pop-up windows), Figure 2.

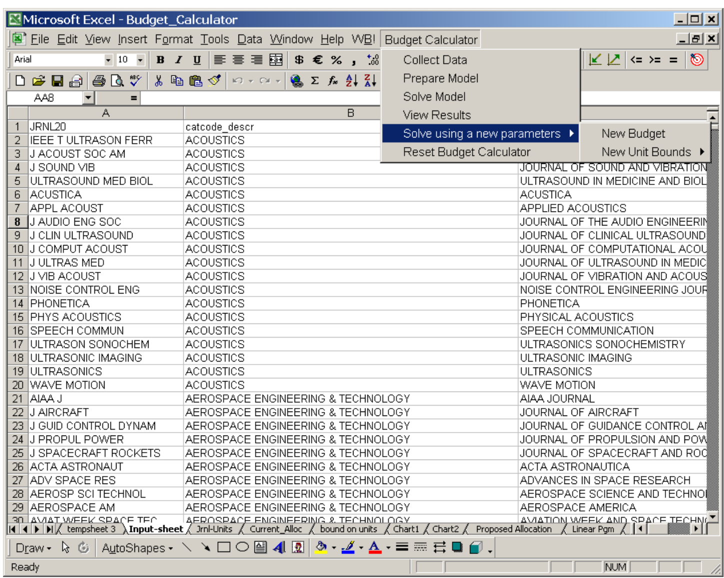

Figure 2: The DSS graphical user interface

The ISI database is in MS Access. By querying the database, the number of citations and publications and the categories are obtained. This data is then exported to an MS Excel spreadsheet. In addition, the user has to input information on the locations of the currently held journals in the library units, the previous year allocation, and the number of currently held journals in each category, which serves as a baseline.

After inputting the data, a series of macros are then executed to create subcategories, forecast usages, create variables and calculate their objective coefficients and average costs. The user is prompted to input the lower and upper bounds on allocation for each of the library units, as well as the total budget figure. Once the system analyzes the information from the data sheets as well as allocation bounds imposed by the user, it creates the model. The solver used is *What's Best 5.0*, Lindo Systems Inc. [9], which is a commercial package for solving mathematical models in MS Excel. Once the solver provides a solution, the system evaluates the results and reports how many journals should be purchased in each category and how much monies should be allocated to each category and each library unit. In addition, it also generates a comparison between the current allocation and the proposed allocation. The results are summarized in numerical figures on a spreadsheet and the system also creates charts and graphs to depict the allocation to various categories and library units. Figure 3 gives a snapshot of the system wherein it shows how the available funds ($2.7 million in this example) should be distributed among the library units.

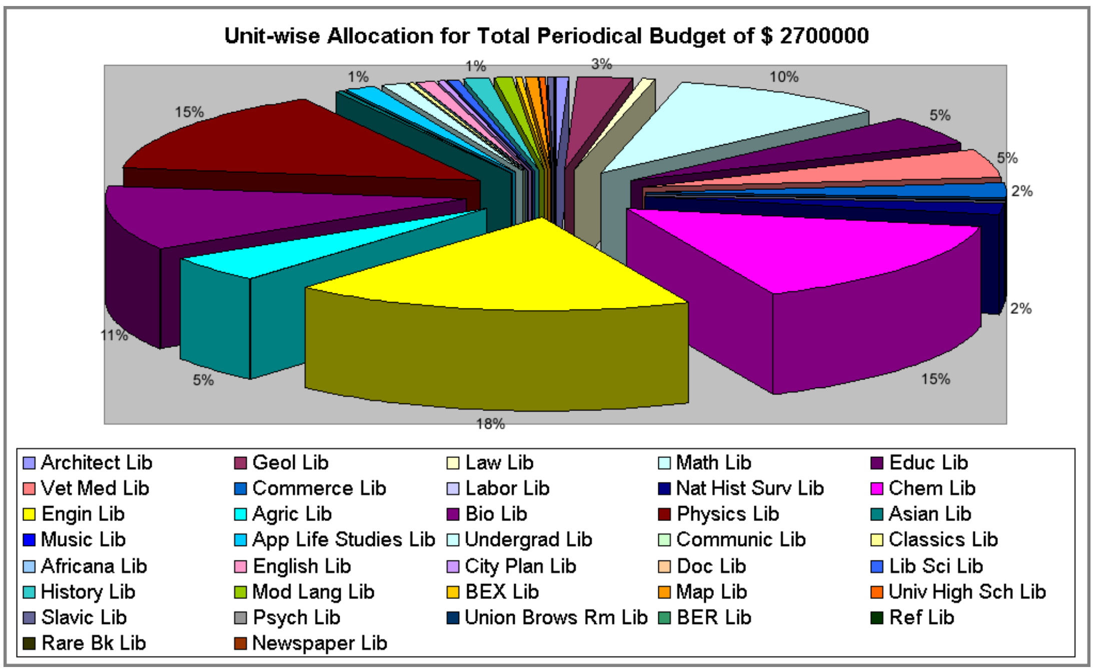

Figure 3: A pie-chart depicting allocation of funds to the library units

The decision maker can generate multiple allocations by resetting upper and lower bounds on the amount of funds that can be allocated to each library unit. He/she has also the option of changing the budget figure and/or the weights.

3.1 Computational Results

The DSS is tested on the data available from the UIUC Library. 37 library units and an annual periodical budget of $2.7 million are considered. The ISI database contains approximately 6000 journals that are classified into 226 categories, which yields a model with 2062 unknown variables and 1894 constraints. A sample of the results is tabulated in Table 1.

For approximately 80% of the categories the solution obtained from the DSS differs from the baseline by less than 5 journals, which complies quite well with the current scenario. For some categories the

Category	Number of Journals Proposed by the Model	Number of Journals Currently Purchased	Difference
Agriculture	85	82	3
Behavioral Sciences	31	31	0
Chemistry, Organic	12	37	-25
Classics	20	25	-5
Communication	32	31	1
Economics	135	135	0
Engineering, Electrical and Electronics	139	144	-5
Ergonomics	7	6	1
General and Internal Medicine	72	42	30
Geography	42	40	2
History	103	105	-2
Language and Linguistics	80	80	0
Management	50	45	5
Neuroscience	86	99	-13
Operations Research and Management Science	37	33	4
Optics	37	32	5
Plant Sciences	102	91	11
Sociology	78	81	-3
Veterinary Sciences	109	97	12
Zoology	78	78	0

Table 1: A sample of results

model suggests significant changes. An in-depth analysis of the data available for these categories revealed the following: *Organic Chemistry* represents a typical category where the cost of journals is relatively high (average price is approximately $2400), whereas the UIUC faculty have published in and cited only a handful of journals in this category. Therefore the model takes into account the balance between the price and the usage and proposes to purchase a fewer number of journals in this category. The other extreme case is represented by categories such as *General and Internal Medicine*, where the model proposes to buy more journals than are being currently purchased. Such categories usually have journals which are relatively cheap (average price of journals in *General and Internal Medicine* is approximately $200) and/or have a large number of publications and citations by the UIUC faculty. Discussions with librarians revealed that in most cases where the model proposed to increase spending for a category, there had already been such requests and recommendations under consideration within the library. All fields pertaining to health and medicine is one such example, where there has been a growing concern to increase spending in these areas and the model clearly corroborates it. While the librarians were satisfied with the results at large and aware of deficits in most of the categories in which the DSS proposes to increase the existing holdings, the situation is somewhat different in categories where a significant decrease is proposed. These subject areas are well established on campus and they have never contemplated the balance between the usage and price. Running the model with an infinite budget (i.e. not taking into account the journal prices) reveals that the new solution no longer differs from the baseline in these categories. This shows that the high journal price is a driving factor in these categories and that based on the model output it is hard to justify these high prices.

The results reveal some features of this model. First, it helps in identifying subject areas which require more attention by the decision makers. These subject areas are those where the proposed allocation is incongruent with the current allocation. It is quite easy to trace back the results to the input data and reason out why a particular subject area deserves the proposed allocation. Second, since the allocations can be generated without the use of weights that prioritize the categories, it rules out any subjectiveness and bias in the allocation.

4 Conclusions and Future Research

The authors believe that the proposed model can fit well with any academic library, which has multiple units and a single collections budget for all of the units. This model is usage based and different libraries are likely to use different measures of usage. While deciding upon a suitable measure of usage, it is essential to verify that it is well correlated with the historical usage data. Usage can be measured indirectly by combining several factors as in the case of the UIUC Library. It can also be measured directly, e.g. a library where reshelving procedures are well established and integrated within an information system could use this data to directly measure its usage of materials. The authors foresee that such a model can be used to facilitate the budget allocation process for a library and it can be used consistently year after year due to its ability to capture trends in usage. The ease of generating allocations quickly by changing a few parameters in the model makes it a worthwhile tool for librarians, who can compare different allocations and evaluate them for decision making purposes.

A similar approach can be developed for monographs. Monographs can be classified into categories using systems such as dewey decimal numbers. The usage in this case is measured as a function of both the circulation in a category and the life-cycle of the materials. The latter takes into consideration the "depth of collections". The depth refers to the need to acquire materials primarily for the reason of collecting rare material and material purchased in anticipation of long-term future use. Such material is expected to have very marginal usage at present and therefore has to be considered differently. The authors believe that the depth of collections can be measured by analyzing a life span of borrowed monographs in different disciplines.

Over the past few decades, as computer and information technology have grown rapidly, libraries have started considering web resources and electronic materials as a significant and invaluable element of their collections. Clearly, with the advantages electronic resources have to offer (such as availability of information even in remote locations, substantially reduced storage space and low cost) it is expected that popularity of this new medium will continue to escalate. Most of the popular journals can now be accessed online with appropriate permissions from the publisher and many conference proceedings are now available only on CDs. In a recent report, The Association of Research Libraries [2] estimated that an average library spent 12.9% of its budget on electronic materials in the fiscal year 1999-2000. This is almost a fourfold increase from a mere 3.6% in fiscal year 1992-1993. Such an increase necessitates a systematic and well thought plan to allocate these monies. The second half of the previous decade has witnessed a surfacing concern among libraries, publishers and other organizations to measure the usage of electronic resources and collect usage statistics. The ARL's E-metrics project[1] is one such noteworthy effort in this context, which deals with defining and collecting data on usage of electronic resources. In addition, the International Coalition of Library Consortia (ICOLC) lays down guidelines for data-collection policies.

In the E-metrics project they classify electronic material in two categories.

1. Material accessible via the internet on subscription basis. An example of this are online journals.

2. Electronic material available within the physical realm of the library. This includes storage media such as CDs, DVDs, E-books, videotapes, audiotapes, etc.

The E-metrics project has accomplished an in-depth study on how to capture, report and analyze usage statistics on materials of these types. A basic idea for measuring the usage of materials of the former type requires setting up an information system where users are required to log on via a dedicated web domain to gain access to materials on a publisher's website. Such a system can keep track of usage of materials

[1] http://www.arl.org/stats/newmeas/emetrics

accessed and publishers can record this information and feed it back to the library (or university). A possible measure of usage in this case can be the number of times a particular electronic journal is accessed by users in a given period of time. If a separate budget for electronic material is being held, then an identical model can be used by appropriately defining the usage values. If a library has a single budget for both the electronic material and traditional periodicals, then the usage values for online periodicals can be calculated by studying the correlation in a category between the number of citations in an online periodical and the number of users' accesses to the periodical. The possible approach is very similar to the approach we use to combine the number of citations and publications (see Section 2.1.1).

For keeping track of usage of materials of the second type they discuss networked environments where a server can be used to keep track of how many times a particular CD or any other electronic resource is accessed at a terminal. In case it is also possible to record who is accessing the material[2], user profiling can lead to information on what materials and to what extent they are being used by the academic departments. If such data is available for each department, regression based forecasting (discussed in Section 2.2) would be more applicable and it would yield a more appropriate usage measure as compared to that based on time-series forecasting.

Acknowledgements

The authors are grateful to Professor William H. Mischo and Professor Alexander Scheeline for their priceless support and cooperation, without which this work would not have been possible. In addition they are thankful to the UIUC Library for providing the financial support.

References

[1] ALBRIGHT, S. C. (2001). *VBA for Modelers: Developing Decision Support Systems with Microsoft Excel*. Duxbury Thompson Learning.

[2] ASSOCIATION OF RESEARCH LIBRARIES . www.arl.org.

[3] BUDD, J. (1991). Allocation formulas in the literature: A review. *Library Acquisitions: Practice and Theory* **15**, 95–107.

[4] E. BOURGEOIS, P. COHEN, J. D. and NATESAN, C. (1998). Faculty-determined allocation formula at Southwest Texas State University. *Collection Management* **23**, 113–123.

[5] GERMAN, L. B. and SCHMIDT, K. A. (2001). Finding the right balance: campus involvement in the collections allocation process. *Library Collections, Acquisitions & Technical Services* **25**, 421–433.

[6] GLEESON, M. and OTTENSMANN, J. (1994). A decision support system for acquisitions budgeting in public libraries. *Interfaces* **24**, 107–117.

[7] ISI THOMPSON SCIENTIFIC . www.isinet.com.

[8] KEAN, G. (2001). 14th annual study of journal prices for scientific and medical society journals. *The Newsletter for Journal Publishers, Allen Press Inc.*

[9] LINDO SYSTEMS INC. (2000). *What's Best, User's manual*.

[2] See the E-metrics project on the issues involved with recording the identity.

[10] LOWRY, C. B. (1992). Reconciling pragmatism, equity, and need in the formula allocation of book and serial funds. *College and Research Libraries* **53**, 121–138.

[11] MAKRIDAKIS, S. and WHEELWRIGHT, S. (1989). *Forecasting Methods for Management*. John Wiley and Sons.

[12] M. NIEMEYER, L. LAWSON, G. M. and SLATTERY, C. (1993). Balancing act for library materials budgets: Use of a formula allocation. *Technical Services Quarterly* **11**, 43–60.

[13] PROMIS, P. (1996). Developing a databased budget allocation strategy: The university of arizona library experience. *Collection Building* **15**, 5–9.

[14] REPORT OF THE LIBRARY ALLOCATION STEERING COMMITTEE (2001). University of Illinois at Urbana-Champaign. http://www.provost.uiuc.edu/provost/committees/libraryallocation.html.

[15] WISE, K. and PERUSHEK, D. (1996). Linear goal programming for academic library acquisitions allocations. *Library Acquisitions: Practice and Theory* **20**, 311–327.

[16] ZILLA, S. (1984). A network optimization model for budget allocation in a multi-campus university. *Journal of Operations Research Society* **35**, 749–757.

Appendix

Here we formally define the model. The following variables are defined.

1. x_i variables : For every subcategory i, x_i is the number of journals purchased in subcategory i from among all the journals that are not cross-listed.

2. y_{ij} variables : For every subcategory i and j, y_{ij} is the number of journals purchased from the set of all those journals that are cross-listed in subcategories i and j only.

3. z_k variables : For every journal k that is cross-listed more than twice, z_k is 1 if the journal should be purchased and 0 otherwise.

The x and y variables are nonnegative integers and the z variables are binary.

The model maximizes the usage subject to the availability of funds and bounds on allocations to the library units. The following quantities are needed.

S	the set of all subcategories
L	the set of all library units
b	the budget for the forthcoming fiscal year
K	the set of all journals that are cross-listed in more than two subcategories
$\tilde{K}_i$	the set of all journals in K and subcategory i
$\hat{K}_l$	the set of all journals in K that are (or would be) in library unit l
$\tilde{c}_i$	the average price of journals among all non-cross-listed journals in subcategory i
$\bar{c}_{ij}$	the average price of journals among all the journals cross-listed exactly in subcategories i and j
$\hat{c}_k$	the price of a journal $k \in K$
n_i	the number of journals in subcategory i
p_i	the number of journals in subcategory i that are not cross-listed
q_{ij}	the number of journals cross-listed exactly in subcategories, i and j
$\underline{m}_l$	the minimum amount of funds that is allowed to be allocated to library unit l
$\overline{m}_l$	the maximum amount of funds that is allowed to be allocated to library unit l
$\tilde{f}_{il}$	the fraction of journals in subcategory i that are held in library unit l
$\bar{f}_{ijl}$	the fraction of journals cross-listed exactly in subcategories i and j that are held in library unit l

The values $\underline{m}_l$ and $\overline{m}_l$ are assigned by a decision maker. In addition, let w_i be the weight assigned to category i. These weights are also assigned by the decision maker and they reflects university priorities in given areas. It is assumed that $\sum_{i \in S} w_i = 1, w_i \geq 0$ for all $i \in S$.

Let u_i be the usage values per journal of subcategory i, which is calculated as follows

$$u_i = \frac{F_{Ti}}{n_i}$$

The objective coefficients $\tilde{u}_i$, $\bar{u}_{ij}$, and $\hat{u}_k$ (usage coefficients of x,y and z variables respectively) are defined in the objective function of the model as follows.

$$\begin{aligned}
\tilde{u}_i &= w_i u_i && \text{for all } i \in S \\
\bar{u}_{ij} &= \frac{w_i u_i + w_j u_j}{2} && \text{for all } i \in S, j \in S \\
\hat{u}_k &= \frac{\sum_{i \in S_k} w_i u_i}{|S_k|} && \text{for all } k \in K
\end{aligned}$$

Here for $k \in K$, $S_k \subseteq S$ denotes the set of all subcategories that the journal k is assigned to. In the definition of $\bar{u}$ and $\hat{u}$ the mean usage among all the subcategories to which a journal is assigned is considered.

The model reads

$$\max \sum_{i\in S} \tilde{u}_i x_i + \sum_{i\in S, j\in S} \bar{u}_{ij} y_{ij} + \sum_{k\in K} \hat{u}_k z_k$$

$$\sum_{i\in S} \tilde{c}_i x_i + \sum_{i\in S, j\in S} \bar{c}_{ij} y_{ij} + \sum_{k\in K} \hat{c}_k z_k \leq b \tag{1}$$

$$x_i + \sum_{j\in S} y_{ij} + \sum_{k\in \tilde{K}_i} z_k \leq n_i \qquad \text{for all } i \in S \tag{2}$$

$$x_i \leq p_i \qquad \text{for all } i \in S \tag{3}$$

$$y_{ij} \leq q_{ij} \qquad \text{for all } i \in S, j \in S \tag{4}$$

$$\underline{m}_l \leq \sum_{i\in S} \tilde{f}_{il} \tilde{c}_i x_i + \sum_{i\in S, j\in S} \bar{f}_{ijl} \bar{c}_{ij} y_{ij} + \sum_{k\in \hat{K}_l} \hat{c}_k z_k \leq \overline{m}_l \qquad \text{for all } l \in L \tag{5}$$

x nonnegative integer, y nonnegative integer, z binary .

The objective function maximizes the usage in the future time period. (1) assures that the model does not allocate more funds than the available budget. The set of constraints (2) guarantee that no more journals than the maximum available in each subcategory are allocated. (3) and (4) impose similar bounds on non cross-listed and cross-listed journals. Constraints (5) approximate that the model does not allocate to a library unit more funds than the given lower and upper bound.

Additional information usually sought by librarians can easily be found by slightly adjusting the model. For example, the number of journals that should be ideally purchased in each category can be derived by assuming that $\underline{m}_l = 0, \bar{m}_l = \infty$ for each $l \in L$. The computed number of journals in each category is the sum of the left hand sides of (2) corresponding to the two subcategories arising from the category. A second question is to determine the number of journals that should be purchased in each category regardless of journal prices, i.e. taking into account only the usage. This can be obtained by setting $\underline{m}_l = 0, \bar{m}_l = \infty$ for each $l \in L$ and $b = \infty$. Note that the model takes into account the balance between the journal prices and the usage.