Title
The 2nd POMS Applied Research Challenge 2016 Awards

Journal
Production and Operations Management, 25(12)

ISSN
1059-1478

Authors
Caro, Felipe
Tang, Christopher S

Publication Date
2016-12-01

DOI
10.1111/poms.12637

Peer reviewed

POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS
Vol. 25, No. 12, December 2016, pp. 2002–2013
ISSN 1059-1478 | EISSN 1937-5956 | 16 | 2512 | 2002
DOI 10.1111/poms.12637

The 2nd POMS Applied Research Challenge 2016 Awards

Felipe Caro, Christopher S. Tang
UCLA Anderson School of Management, UCLA, 110 Westwood Plaza, Los Angeles, California 90095, USA,
felipe.caro@anderson.ucla.edu, chris.tang@anderson.ucla.edu

The 2nd POMS Applied Research Challenge (POMS ARC) took place at the annual POMS meeting in Orlando. The POMS ARC is a bi-annual initiative was launched in October 2012. Its objective is to encourage POMS members – faculty and students – to conduct rigorous applied research that is relevant and innovative. Applied research is understood as academic work that meets the following requirements:

1. The work involves a real operation;
2. The problem is new or not well-solved;
3. The problem requires some innovative ideas to solve it;
4. The study makes a convincing case of its relevance to practice; and
5. The study could eventually be published in a journal like POM or similar.

The 2nd POMS ARC was chaired by Felipe Caro (UCLA Anderson School of Management) with the support of past POMS President Christopher Tang (UCLA Anderson School of Management). The POMS ARC had two review panels: the Practitioner Judge Panel formed by distinguished POMS practitioners and the Academic Panel formed by academics with a track record in applied research. The members of the Practitioner Judge Panel were: Corey Billington (e3 Associates), Srinivas Bollapragada (GE Global Research), Edwin Keh (HKRITA, former Walmart COO), Dino Petrarolo (CCI Inc.). The members of the Academic Panel were: Feryal Erhun (Cambridge Judge Business School), Nagesh Gavirneni (Cornell University), ManMohan Sodhi (City University London), and Felipe Caro (UCLA).

There were more than 35 paper submissions, which went through a two-step review process. First, the Academic Panel selected three finalists. Then, the finalists were invited to present their work at the 2016 POMS Annual Meeting Conference in Orlando where the Practitioner Judge Panel selected the winner based on the following criteria: (i) financial benefit; (ii) scalability to other industries and domains; (iii) managerial usability; and (iv) elegance, clarity, and depth of insight. All the finalists will publish an extended abstract of their work in the POM journal. The first prize was accompanied by a $2000 honorarium.

The 2016 prize winners of the 2nd POMS Applied Research Challenge are as follows:

First Prize

Optimal Purification Decisions for Engineer-to-Order Proteins at Aldevron

Tugce Martagan, Eindhoven University of Technology
Ananth Krishnamurthy, University of Wisconsin-Madison
Peter A. Leland, Aldevron
Christos T. Maravelias, University of Wisconsin-Madison

Finalists

From Predictive to Prescriptive Analytics

Dimitris Bertsimas, MIT
Nathan Kallus, Cornell Tech
Amjad Hussain, CEO Silkroute

Quantifying Uncertainties Using Expert Assessments in a Dynamic New Product Development Environment

Saurabh Bansal, Penn State
Genaro J. Gutierrez, UT Austin
John R. Keiser, Dow AgroSciences

The extended abstracts of the first prize winning and two finalist papers are provided next.

Optimal Purification Decisions for Engineer-to-Order Proteins at Aldevron

Tugce Martagan
School of Industrial Engineering, Eindhoven University of Technology, De Rondom 70, Pav. E17, Eindhoven, 5612AZ Netherlands, t.g.martagan@tue.nl

Ananth Krishnamurthy
Department of Industrial and Systems Engineering, University of Wisconsin-Madison, 1513 University Avenue, 3121C Mechanical Engineering Building, Madison, WI 53706, USA, ananth@engr.wisc.edu

Peter A. Leland
Aldevron, 5602 Research Park Blvd, Madison, WI 53719, USA, leland@aldevron.com

Christos T. Maravelias
Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, 2004 Engineering Hall, Madison, WI 53706, USA, christos.maravelias@wisc.edu

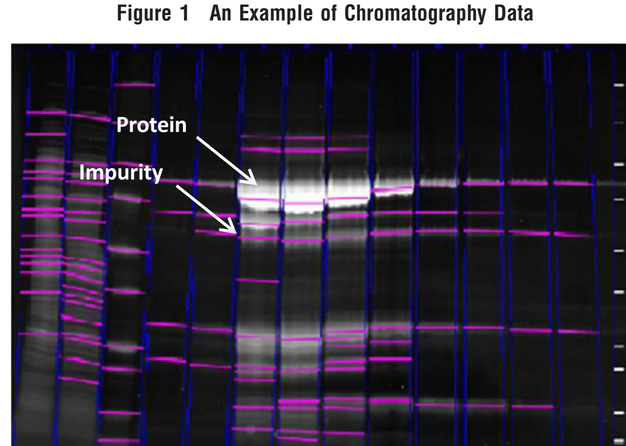
1. Introduction

We investigate protein purification operations in pharmaceutical research and development. Each production order represents an engineer-to-order protein that needs to be purified using chromatographic separation. An order has predetermined purity and yield requirements, and a biomanufacturer incurs high penalty costs when these requirements are not satisfied. However, achieving these requirements is often challenging because of an inherent trade-off between the batch purity and protein yield. In this setting, it is of practical importance to answer the following research questions: (i) For a given starting material, is it possible to determine whether the final purity and yield requirements specified by the customer are achievable at all? (ii) How to control the purification operations to achieve the maximum expected profit? To answer these research questions, we develop a dynamic optimization framework in close collaboration with Aldevron, a contract biomanufacturer specializing in protein manufacturing.

2. Background in Protein Purification

Chromatography is a key technique used in protein purification (Farid 2009, Polykarpou et al. 2011). The objective of a chromatography operation is to separate the protein of interest from unwanted impurities in order to meet a predetermined *purity requirement* specified by the end user or application. The purity represents the fraction of the amount of protein of interest available in a batch. Figure 1 presents an example of chromatography output. In Figure 1, the columns on the x-axis represent *lanes* and can be thought as equivalent to a discrete time interval. Each lane is comprised of some amount of the protein of interest as well as some amount of the unwanted impurity. The y-axis in Figure 1 represents the molecular mass of the protein and impurity in each lane. The scientist performing the purification must decide which lane to *"pool."* For example, if she pools the lanes 4–3, then she collects a large amount of protein at the expense of large amount of impurity. On the other hand, if she pools the lanes 6–8, she compromises on the protein yield but improves the purity. This illustrates the purity-yield trade-off encountered

Figure 1 An Example of Chromatography Data



[Correction added on 14 November 2016, after first online publication: the authors' affiliations have been corrected accordingly.]

in chromatography operations. In a typical protein purification process, the first step consists of *scouting runs* where the scientist collects chromatography data. Next, the scientist performs *validation runs* to mitigate the risks, and then proceeds with larger scale *production runs*.

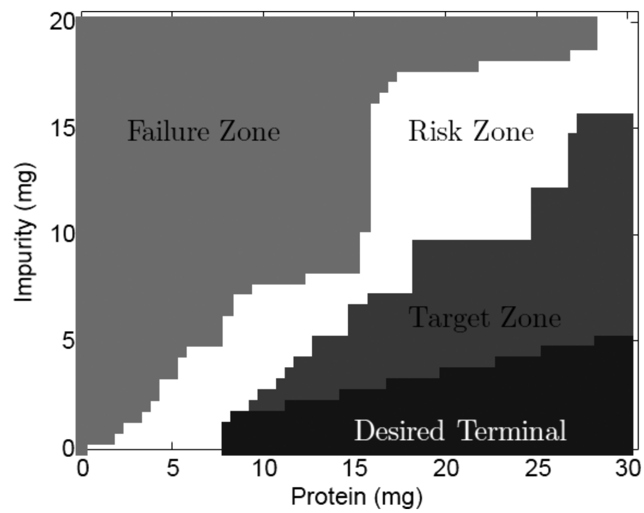
The key challenges in choosing the pooling windows can be summarized as follows: (1) *Production requirements.* Each order has predetermined purity and yield requirements specified by the end user or application. However, achieving these requirements is often challenging in practice due to the purity-yield trade-off involved in chromatography operations. (2) *Engineered proteins.* Each order is manufactured for the first time as part of an R&D project. (3) *Uncertainty.* The outcome of a chromatography step involves uncertainty in purity and yield. (4) *Interlinked decisions.* Purification involves multiple chromatography steps in series. (5) *Variability in the starting batch.* The starting material involves high variability in terms of the amount of protein and impurity.

3. The Model, Structural Analysis, and Insights

We model the purification problem as a discrete-time Markov decision process where each decision epoch represents the beginning of a chromatography step. The state denotes the amount of protein and impurity in the batch. The action space consists of the set of pooling windows associated with each chromatography step. In addition, the scientist has the option for stopping the purification process at the beginning of a chromatography step. The underlying state transitions are captured using real-world scouting data. The biomanufacturer incurs large failure costs if the final batch does not comply with the purity requirement. In addition, yield penalty costs are incurred for each unit of protein in short. The objective is to maximize the total expected profit through identifying the optimal pooling windows and the optimal stopping time for a particular purification project.

We analyze the structural properties of the optimization model, and establish theoretical results that provide guidelines for practitioners. Our analysis partitions the state space into distinct subsets called *decision zones*. These decision zones are namely the *failure zone*, *risk zone*, and *target zone*. Figure 2 illustrates the decision zones associated with the first chromatography step of a particular purification project at Aldevron. The decision zones provide an objective assessment of the starting material, manufacturing capabilities and failure risks at the beginning of each chromatography step. Using the insights obtained from the decision zones, we then provide practical guidelines to maximize the expected profit of a protein purification project. The proposed zone-based decision making approach is particularly easy to implement and use in practice.

Figure 2 Decision Zones for a Particular Order at Aldevron



3.1. Failure Zone

The failure zone is a subset of the state space (i.e., protein and impurity amounts) where the biomanufacturer has no financial incentives for performing the purification project. In other words, a starting batch belonging to the failure zone represents a “bad” material that would eventually lead to a substantial financial loss for the biomanufacturer.

3.2. Target Zone

The target zone is generated using the worst-case analysis, and represents the “good” starting material that eventually leads to a successful purification project with certainty. More specifically, if the starting material is within the target zone, then the biomanufacturer can provide performance guarantees to its client in terms of achieving both the yield and purity requirements. The target zone provides important insights for practitioners. For example, it establishes performance guarantees for achieving the production requirements. Such performance guarantees lead to a competitive advantage by ensuring customer satisfaction despite the manufacturing challenges. These performance guarantees also provide significant visibility to the production pipeline.

3.3. Risk Zone

We define the risk zone as all states that are neither in the target zone nor in the failure zone at the beginning of a chromatography step. This subset of the state space is a measure of the financial risks associated with purifying a particular batch. For example, if a batch is in the risk zone, it could either achieve the final purity and yield requirements or fail to do so.

The *desired terminal* zone illustrated in Figure 2 represents all the protein and impurity pairs that already satisfy the production requirement associated with a particular order.

3.4. Optimal Policies

We characterize the optimal purification policies at each chromatography step in terms of the decision zones described above. For example, it is optimal to stop the purification project if the starting material is an element of the failure zone. In this case, the biomanufacturer is financially better off with not accepting the order (i.e., failing earlier than later). On the other hand, if the starting material is within the target zone, we show that the necessary condition for the optimal policy is to perform the purification in a way that the process stays within the target zone of the next chromatography step. If the starting material is within the risk zone, then we show that it is optimal to perform the purification in a way that the purity requirement is always met with the least possible yield loss. We also prove that the decision zones have a threshold type structure which can be easily adopted in practice. The insights obtained from the decision zones and optimal polices are then used to generate a state aggregation and action elimination procedure to solve industry-size problems.

4. Implementation Results and Conclusions

Since the implementation of the optimization framework, Aldevron has realized an average of 25% reduction in lead times and 20% reduction in costs due to the following factors:

1. *Formal assessment of the risks and better understanding of manufacturing capabilities.* The optimization framework provides a formal assessment of the business risks and financial trade-offs involved in protein purification operations. The proposed zone-based decision making approach enables a quick and reliable analysis of the manufacturing capabilities leading to better and easier communication with the clients. The knowledge on “guaranteed performance” or “guaranteed failure” obtained by the end of scouting runs provides significant visibility in the production pipeline.
2. *Reduction in the number of validation runs.* The optimization framework allowed reducing the number of validation runs needed prior to full scale production. The process data obtained from the scouting runs is used as an input of the optimization model to generate the optimal policies.
3. *Process economics taken into consideration.* Prior to the use of the optimization framework, potential operating policies were assessed based on historical experience. As a result, the scientists used to focus on the underlying biology and chemistry of these processes, and could not fully capture the business risks and financial implications of the pooling decisions. In contrast, the optimization model provides a formal framework that considers the financial risks and trade-offs involved in protein purification operations.

To facilitate the implementation of the proposed optimization framework, a decision support tool has been developed at Java. The tool provides a user-friendly interface to generate the decision zones and the optimal policies in practice. Feedback from the broader biomanufacturing community beyond Aldevron has also been a critical part of the problem definition, analysis and validation. For example, we organized a series of working group sessions with the local biomanufacturing companies during various phases of this research (BioWGS 2014, Foti et al. 2016, Martagan et al. 2014, 2016). Applications of operations research techniques are mostly new to the biomanufacturing community. As more companies like Aldevron embrace operations research and integrate it into practice, we believe that regulatory authorities might mandate the use of such approaches to reduce costs and lead times in the biomanufacturing research and development.

References

BioWGS. 2014. Biomanufacturing working group sessions. Organized by the University of Wisconsin-Madison and Aldevron. Available at http://www.bioforward.org/event/id/392756/Challenges-and-Opportunities-in-Managing-Biomanufacturing-Operations.htm (accessed date December 12, 2014).

Farid, S. 2009. Process economic drivers in industrial monoclonial antibody manufacture. U. Gottschalk, ed. *Process Scale Purification of Antibodies*. John Wiley & Sons, Hoboken, NJ, 239–261.

Foti, T., A. Krishnamurthy, P. Leland. 2016. Aldevron collaborates with the University of Wisconsin to develop decision support tools for biomanufacturing. *BioForward Speaker Series*.

Martagan, T., A. Krishanmurthy, P. Leland, T. Foti. 2014. Decision support tools to reduce lead times and costs in biomanufacturing, Bioscience Vision Summit. Available at http://www.bioforward.org/general/custom.asp?page=Biomanufacturing (accessed date January 25, 2015).

Martagan, T., A. Krishnamurthy, C. T. Maravelias. 2016. Optimal condition-based harvesting policies for biomanufacturing operations with failure risks. *IIE Trans*. **48**(5): 440–461.

Polykarpou, E. M., P. A. Dalby, L. G. Papageorgiou. 2011. Optimal synthesis of chromatographic trains for downstream protein processing. *Biotechnol. Prog*. **27**(6): 1653–1660.

Production and Operations Management 25(12), pp. 2002–2013, © 2016 Production and Operations Management Society

Inventory Management in the Era of Big Data

Dimitris Bertsimas
MIT Sloan School of Management and Operations Research Center, 77 Massachusetts Ave., E40-111, Cambridge, MA 02139, USA, dbertsim@mit.edu

Nathan Kallus
Cornell University and Cornell Tech, 111 Eighth Ave #302 New York, NY 10011, USA, kallus@cornell.edu

Amjad Hussain
Silkroute, 950 Stephenson Hwy Troy, Michigan 48083 USA, amjad.hussain@silkrouteglobal.com

1. Introduction

The explosion in the availability and accessibility of machine-readable data is creating new opportunities for better decision making in applications of operations management. The swell of data and advances in machine learning have enabled applications that predict, for example, consumer demand for video games based on online web-search queries (Choi and Varian 2012) or box-office ticket demand based on Twitter chatter (Asur and Huberman 2010). In the context of inventory management, demand is the key uncertainty affecting decisions and such works suggest a potential opportunity to leverage large-scale web data to improve inventory decisions, for example, for stocking video game titles or allocating cinemas of varying capacities. There are also many other applications of machine learning, including Da et al. (2011), Goel et al. (2010), Gruhl et al. (2004, 2005), Kallus (2014), that use large-scale and web-based data to generate predictions of quantities that may in fact be of interest in operations management applications. By and large, however, these applications and the machine learning techniques employed do *not* address optimal decision-making under uncertainty that is appropriate for operations management problems and, in particular, for inventory management.

We study how these data, leveraged appropriately, can correctly and successfully inform inventory management decisions and provide a competitive edge. We focus on a particular case study of the distribution and manufacturing arm of a global media conglomerate (henceforth, the vendor), which, as a distributor of multi-media, is among the three largest in the world. The vendor, which shall remain unnamed, is a direct customer of Silkroute, a provider of analytics platforms for managing manufacturing, distribution, and retail operations. The vendor, which ships an average of 1 billion units in a year, as well as the media retail industry at large, is under increased pressure to improve operations and lower costs in the face of increasing digitalization, declining sales, and diminishing shelf space. The heightened importance and consequence of good inventory decisions provide an excellent case study of the use of large-scale data for achieving a competitive edge in a squeezed industry.

We consider the vendor's VMI (vendor-manage inventory) operations in selling over half-a-million entertainment titles on CD, DVD, and BluRay at major European retailers with over 20,000 locations. To inform VMI decisions, we leverage transactional records collected and organized by the Silkroute platform, data we harvested from public Internet sources including IMDb.com (International Movie Database) and RottenTomatoes.com, and search query volume data provided by Google Trends.

To leverage these data, we employ recent data-driven optimization techniques developed by Bertsimas and Kallus (2014) that address the *conditional stochastic optimization problem*:

$$\begin{aligned} v^*(x) &= min_{z\epsilon\mathcal{Z}}\mathbb{E}[c(z;Y)|X=x], \\ z^*(x) &\in argmin_{z\epsilon\mathcal{Z}}\mathbb{E}[c(z;Y)|X=x], \end{aligned} \tag{1}$$

wherein, on the basis of an observation of auxiliary covariates $X \in \mathbb{R}^{d_x}$, a decision $z(x)$, constrained in a feasible space $\mathcal{Z} \subset \mathbb{R}^{d_z}$, is chosen in an optimal manner to minimize an uncertain cost $c(z;\ Y)$ that depends on a random variable $Y \in \mathbb{R}^{d_y}$. For example, in the context of media retail inventory management, the uncertain quantities Y of direct impact on costs are the demands for stocked products; the decisions are quantities $z \geq 0$ for each product, constrained by limited capacity $1^{\mathrm{T}} z \leq K$; and, the auxiliary covariates X that may help us choose the best quantities may include recent sale figures, volume of Google searches for a products or company, news coverage, or user reviews. The solution $z^*(x)$ to problem (1) represents the full-information optimal decision, which, via full knowledge of the joint distribution of X, Y, leverages the observation $X = x$ to the fullest possible extent in minimizing expected costs. In practice, the underlying joint distribution of X, Y is unknown and we must devise a policy $\hat{z}_N(x)$ based only on data $S_N = \{(x^1, y^1), \ldots, (x^N, y^N)\}$. This learning task was addressed in Bertsimas and Kallus (2014), where new methods for this problem are developed, which have two important properties:

Asymptotic optimality:

$$lim_{N\to\infty}\mathbb{E}[c(\hat{z}_N(x);Y)|\mathrm{X}=x]=v^*(x)$$

for almost everywhere x, almost surely.

Tractability: $\hat{z}_N(x)$ can be computed in polynomial time and oracle calls, and, in many important cases, it is solvable using off-the-shelf optimization solvers.

One of the simplest approaches proposed in Bertsimas and Kallus (2014) is based on k-nearest neighbors (kNN), where we let

$$\hat{z}_N^{kNN}(x) \in \mathrm{argmin}_{z\epsilon\mathcal{Z}}\sum\nolimits_{i\in N_k(x)}c(z;y^i),$$
$$N_k(x)=\{i : x^i \text{ is among the } kNN\text{s to x in the data}\}.$$

Various additional methods are developed in Bertsimas and Kallus (2014). The coefficient of prescriptiveness is defined in Bertsimas and Kallus (2014) as

$$P=\frac{\mathbb{E}[c(\hat{z}_N(x);Y)]-\mathbb{E}[\min_{z\in\mathcal{Z}}c(z;Y)]}{\min_{z\in\mathcal{Z}}\mathbb{E}[c(z;Y)]-\mathbb{E}[\min_{z\in\mathcal{Z}}c(z;Y)]},$$

which unitlessly measures the prescriptive content of the auxiliary data X and the efficacy of the policy $\hat{z}_N(x)$ with respect to operational costs in a manner analogous to coefficient of determination R^2 for prediction. In our case study, the rich, large-scale data collected combined with these advances in data-driven optimization account for an 88% reduction in operational costs as measured by P. That is, our approach, based on the data we collect and the prescriptive algorithms we use, takes reduces 88% of excess costs due to uncertainty – a significant advance in addressing the industry's emerging challenges.

2. Problem Description and Formulation

The retail locations in the VMI network range from electronic home goods stores to supermarkets, gas stations, and convenience stores. Under VMI, what is sold at the locations and its replenishment (which is performed weekly) is managed by the vendor. Procurement is done under scan-based trading (SBT), which means that the vendor owns all inventory until scanned at point-of-sale, at which point the retailer procures the unit and sells to the customer. This means that retailers have no cost of capital in holding the vendor's inventory. The cost of a unit is driven primarily by the fixed cost of content production; manufacturing (pressing) media and delivery costs are secondary. Therefore, maximizing network-wide sell-through is the primary objective of the vendor. The limiting factor is capacity: there is limited shelf space (often limited to an aisle endcap display) and generally no storage. Thus, the main loss incurred in over-stocking a particular product lies in the loss of potential sales of another product that sold out (or was not stocked at all) but could have sold more, and there are many potential products. Apart from the limited shelf space, the other primary difficulty is the high uncertainty inherent in the initial demand for new releases, which, at the same time, drive the most sales.

Let $r = 1, \ldots, R$ index the locations, $t = 1, \ldots, T$ index the replenishment periods, and $j = 1, \ldots, d$ index the products. Denote by z_j the order quantity decision for product j, by Y_j the uncertain demand for product j, and by K_r the overall inventory and display capacity at location r. Optimizing sell-through as discussed in the previous paragraph, the problem decomposes on a per-replenishment-period, per-location basis. We therefore wish to solve, for each t and r, the following problem:

$$v^*(x_{tr})=\max\mathbb{E}\left[\sum\nolimits_{j=1}^{d}\min\{Y_j,z_j\}\middle|\mathrm{X}=x_{tr}\right], \qquad (2)$$
$$s.t.\sum\nolimits_{j=1}^{d}z_j\le K_r, z_j\ge 0\ \forall j=1,\ldots,d,$$

where x_{tr} denotes auxiliary data available at the beginning of period t in the $(t, r)^{\text{th}}$ problem.

2.1. Internal Company Data

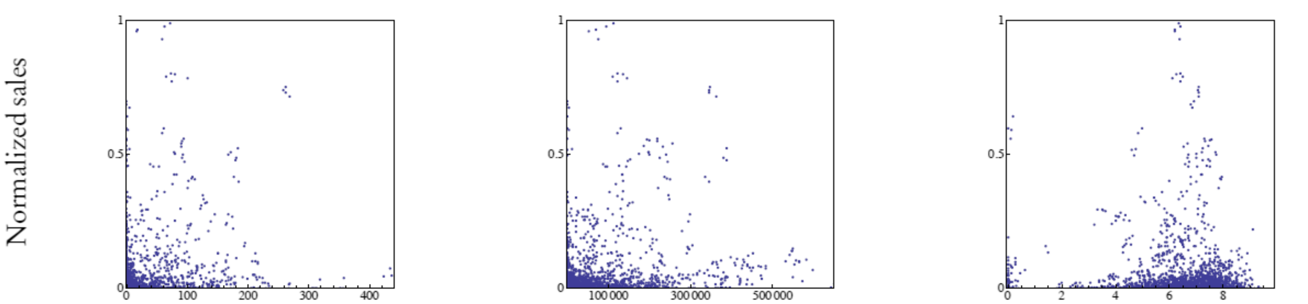
The internal company data collected consists of 4 years of sale and inventory records across the network of retailers, information about each of the locations, and information about each of the items. We aggregate the sales data by week (replenishment period of interest) for each feasible combination of location and item. We use these to collect data on Y,[1] and we include in X the sale volumes of each item at each location over each of the recent 3 weeks (as available; none for new releases), the total sale volume at each location over each of the recent 3 weeks, and the overall mean sale volume at each location over the past year. Information about retail locations includes to which chain a location belongs and the address of the location. We use the Google Geocoding API to parse the address and obtain precise coordinates of the location. We include in X indicators for the country and chain of the location. We also use coordinates to measure search attention as explained below. Information about items include the medium (e.g., DVD or BluRay) and an item title. We disambiguate the item title to obtain a standardized title for the underlying content (e.g., movie name) and use this to collect information about the content as explained below.

Item Metadata, Box Office, and Reviews. To characterize the items and how desirable they may be to consumers, we harvest the data corresponding to each content title on IMDb.com and RottenTomatoes.com (RT). Using data from IMDb, we include in X the number of weeks since the original (e.g., theatrical) release data of the content, content type (film/TV), average user rating, number user ratings, number of awards (e.g., Oscars or Emmys) won and nominated, characteristic vector of first-billed actors' membership in 10 top communities (using Blondel et al. 2008) in the actor-movie graph, indicator vector of closest cluster in a hierarchical clustering of plot summaries by cosine similarity, characteristic vector of reported genres (out of 26), and MPAA rating (if rated). Using data from RT, we include in X the aggregate professional reviewers' score, average user rating, number user ratings, and American box office gross (for films). In Figure 1, we provide scatter plots and correlations of some of these attributes against sale figures in the first week of home entertainment (HE) release.

2.2. Search Engine Attention

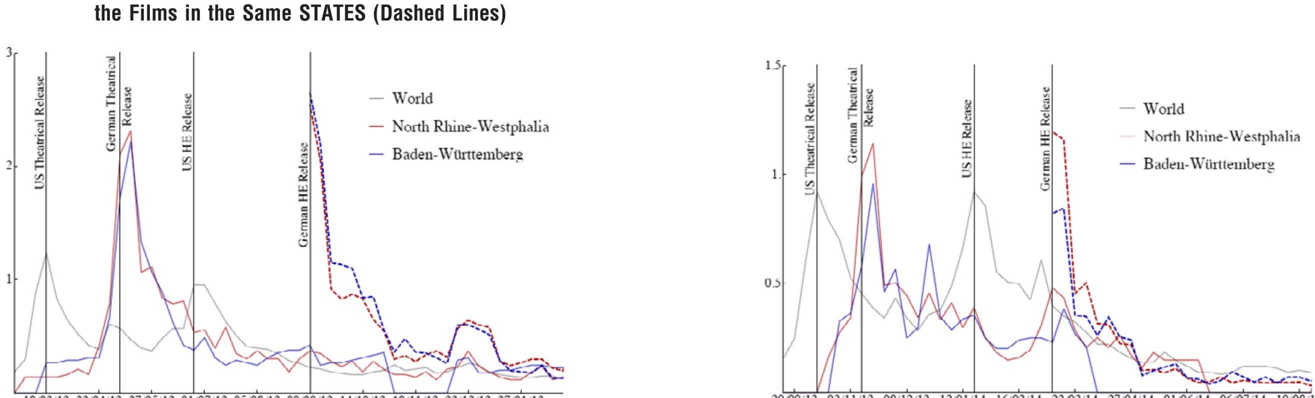
To quantify the attention being given to different titles and to understand the local nature of such attention, we collect search query volume data from Google Trends (GT; www.google.com/trends).[2] GT provides data on the volume of Google searches for a given search term by time and geographic location. For each title, we measure the fraction of Google searches for the search term equal to the original content title in each week from 2011 to 2014 (inclusive) over the whole world, in each European country, and in each country subdivision (states in Germany, cantons in Switzerland, autonomous communities in Spain, etc.). We include in X the total search engine attention to each title over the first two weeks of original release globally, in the country, and in the country-subdivi-

Figure 1 Scatter Plots of Data from IMDb and RT (Horizontal Axes) against Total European Sales during First Week of HE Release (Vertical Axes, Rescaled) and Corresponding Coefficients of Correlation (ρ)



(a) Box office gross (\$M), $\rho = 0.32$ **(b)** Number IMDb votes, $\rho = 0.25$ **(c)** Average IMDb rating, $\rho = 0.02$

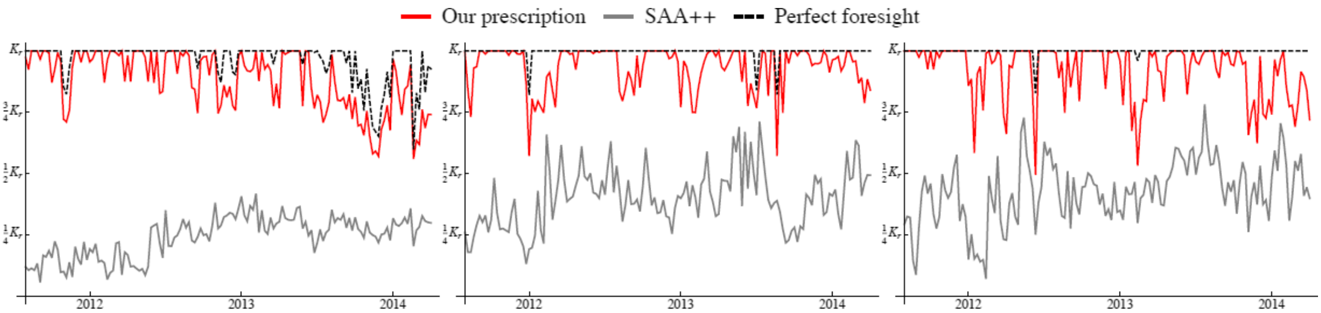
Figure 2 Weekly Search Engine Attention for Two Unnamed Films in the World and in Two German States (Solid Lines) and Weekly HE Sales for the Films in the Same STATES (Dashed Lines)



(a) *Unnamed film 1* **(b)** *Unnamed film 2*

Note. Search engine attention and sales are both shown relative to corresponding overall totals in the respective region. The scales are arbitrary but common between regions and the two plots.

Figure 3 Sell-Through of Various Prescription Over Time



(a) Munich, Germany, $P = 0.89$ **(b)** Paris, France, $P = 0.90$ **(c)** The Hague, Netherlands, $P = 0.86$

sion of each location as well as the search engine attention to each title over each of the recent 3 weeks globally, in the country, and in the country-subdivision of each location. In Figure 2, we compare search engine attention to sales figures in two German states for two unnamed films, which shows the correlation of sales with local *local* search engine attention at original release and the ability of this attention to distinguish sale trends in two locations in the same country.

3. Inventory Prescriptions

Using the data described above, we construct inventory prescriptions $\hat{z}_N(x_{tr})$ for each location r and replenishment period t based on the local weighting approach based on random forest weights (see Bertsimas and Kallus 2014). To evaluate the prescription out-of-sample and as an actual live policy, we consider what we would have done over the 150 weeks from December 19, 2011 to November 9, 2014 (inclusive). At each week, we consider only data from time prior to that week to train the prescription and apply the prescription to the current week. Then, we observe what had actually materialized and score our performance. We compare the performance of our method with the performance of the perfect-forecast policy, which knows future demand exactly (no distributions) and the performance of a data-driven policy without access to the auxiliary data (SAA++).[3] When measured out-of-sample over the 150-week test period, we achieve a coefficient of prescriptiveness $P = 0.88$ averaged over the 20,000 locations, and, in Figure 3, we plot the performance over time at three specific locations. In other words, $P = 0.88$ means that our data X and our prescription $\hat{z}_N(x)$ gets us 88% of the way from the best data-poor decision to the impossible perfect-foresight decision in terms of sell-through volumes.

Notes

[1]In addressing problem (2) in a data-driven context, we face the issue that sales are a censored observation of demand Y. In Bertsimas and Kallus (2014), a remedy is provided in the form of a transformation based on a variant of the Kaplan-Meier method.

[2]Access to massive-scale querying and week-level trends data was generously provided by Google.

[3]For a fair comparison, because demand decay over product lifetime is significant, we let this policy depend on the distributions of product demand based on how long it's been on the market. Due to this handicap we term this policy SAA++.

References

Asur, S., B. Huberman. 2010. Predicting the future with social media. Proceedings of the IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology 492–499.

Bertsimas, D., N. Kallus. 2014. From predictive to prescriptive analytics. arXiv preprint arXiv:1402.5481.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, E. Lefebvre. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech: Theory Exp.* **2008**(10): P10008.

Choi, H., H. Varian. 2012. Predicting the present with google trends. *Econ. Rec.* **88**(s1): 2–9.

Da, Z., J. Engelberg, P. Gao. 2011. In search of attention. *J. Finance* **66**(5): 1461–1499.

Goel, S., J. Hofman, S. Lahaie, D. Pennock, D. Watts. 2010. Predicting consumer behavior with web search. *PNAS* **107**(41): 17486–17490.

Gruhl, D., L. Chavet, D. Gibson, J. Meyer, P. Pattanayak, A. Tomkins, J. Zien. 2004. How to build aWebFountain: An architecture for very large-scale text analytics. *IBM Syst. J.* **43**(1): 64–77.

Gruhl, D., R. Guha, R. Kumar, J. Novak, A. Tomkins. 2005. The predictive power of online chatter. Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. 78–87.

Kallus, N. 2014. Predicting crowd behavior with big public data. Proceedings of the 23rd international conference on world wide web (WWW) companion. 625–630.

Production and Operations Management 25(12), pp. 2002–2013, © 2016 Production and Operations Management Society

Quantifying Uncertainties and Risks Using Managerial Judgments in a Dynamic New Product Development Environment

Saurabh Bansal
Supply Chain and Information Systems, Smeal College of Business, The Pennsylvania State University, University Park, PA 16802, USA, sub32@psu.edu

Genaro J. Gutierrez
Information, Risk, and Operations Management, McCombs School of Business, The University of Texas at Austin, Austin, TX 78705, USA, genaro@austin.utexas.edu

John R. Keiser
Dow AgroSciences, 1562 Taylor Avenue, Marshalltown, IA 50158, USA, JRKeiser@dow.com

1. Research Context: Reliance on Managerial Judgments to Balance Trade-Offs for New Products

The trade-off between too much inventory and too little inventory under uncertain supply and/or demand is an important part of the practice of operations management. The "sweet spot" of inventory, that best balances the overage cost and underage cost, directly depends on the possible scenarios of the uncertainties. These scenarios are succinctly captured in terms of probability distributions that are typically inferred using historical data. But in fast moving industries, firms launch new products frequently and historical data are not always available. A critical hurdle in managing such systems is to estimate the operational uncertainties without data, relying on input from experienced managers or experts. Experienced managers have domain knowledge they can use to provide informed judgments for the probability distributions.

There are two challenges in using expert judgments to deduce probability distributions. First, all judgments provided by an expert are not equally reliable. For example, it is well known that experts cannot provide reliable direct judgments for the standard deviation: the standard deviation is the square root of the second order moment around the unknown mean, as humans we do not have an intuitive feel for this quantity and therefore judgments for the standard deviation tend to be problematic (O'Hagan and Oakley 2004). In contrast, experts find it more intuitive to think about quantiles of distributions (e.g., provide a demand estimate such that there is 50% chance that the demand will be higher than the estimate) and tend to provide better judgments for quantiles as compared to standard deviations. Once these quantile judgments are available, one needs a mechanism to aggregate these quantile judgments and obtain the mean and standard deviation of probability distributions.

The second challenge is that the quantile judgments of even the most experienced or savvy expert are prone to errors. These errors may be present due to an incomplete understanding of business dynamics, or a noisy mental process used to translate the cognitive understanding of the contextual environment into judgmental estimates. Expert-calibration – the quantification of judgmental errors – can be helpful in selecting the optimal aggregation mechanism based on the documented errors. The challenge here is to develop an off-the-shelf approach for this quantification.

Based on these two challenges, our focus in this project was to: (a) Develop an off-the-self approach to quantify errors in experts' quantile judgments, (b) Use this information to obtain estimates of mean and standard deviations from the quantile-judgments, and (c) Quantify the benefit of expert calibration. We first discuss a specific industry application context, followed by the solutions developed and the monetary benefit of using the solutions in the industry application.

[Correction added on 14 November 2016, after first online publication: the authors' affiliations have been corrected accordingly; changes have been made in the equations, font style and grammar for clarity.]

2. Industry Context at Dow AgroSciences

Dow AgroSciences (DAS) is a subsidiary of The Dow Chemical Company and produces seeds for multiple crops, corn being a prominent one. DAS offers a large number of varieties of seed corn to farmers every year. DAS grows these seeds in its fields. The yield, or amount of seed corn obtained per acre of land by DAS during the production of the seeds, is uncertain. Under this yield uncertainty, DAS faces the trade-off between using too much capacity (land) to produce the seed corn which could result in an overage and too little capacity that could result in a shortage. The knowledge of yield distributions is necessary to make this tradeoff using well known mathematical models (e.g., Henig and Gerchak 1990), but the historical data required to make this trade-off are not available barring for a few seeds that have remained farmers' favorite in the last few years. This absence of data is due to the rapid rate of innovation in the commercial seed industry; research scientists continuously develop new varieties of seeds in the laboratory, but there is not sufficient time to produce these seeds repetitively to obtain actual production yield distributions. In the absence of prior data, DAS relies on the judgments of a research scientist who is considered a yield-expert. This scientist has a long experience in hybrid-seed production, and has an intuitive understanding of the biological factors that determine the yields. We used the approach described next to harness his expertise to deduce yield distributions for seeds that are new to DAS' portfolio.

3. Two-Step Approach for Managerial Calibration and Risk Estimation

Our approach has two steps: (i) Calibrate the expert's judgments on a set of distributions for which prior data exist (e.g., seeds that have been sold repetitively in the past), (ii) use the calibration information into an optimization framework to deduce the mean and standard deviation from quantile judgments.

3.1. Step 1: Expert Calibration

In this step, the expert first identifies a set of quantiles that he is comfortable providing his judgments for. In our experience, managers and experts are usually comfortable providing judgments for the median as it has the familiar connotation of 50–50 odds. We also suggest obtaining at least two nonsymmetric quantiles. In our application at DAS, the yield-expert chose to provide the 10th, 50th, and the 75th quantiles. He was used to seeing these quantiles during data analysis on his software and felt confident in providing judgments for these quantiles. Subsequently, the expert considers a number of uncertainties in the same or a closely related context and for which historical data exist. Then he provides his judgments for the selected quantiles for the probability distributions of these uncertainties without looking at the data. At Dow, the yield-expert had a mental model that connected the genetic lineage of seeds to their likely yields, and the expert used this mental model to provide quantile judgments for the yield distributions of seeds with historical data as well the ones without these data.

From this calibration data, the bias and the random noise in the expert's judgments are evaluated. As an example, consider an expert who provides the values of (72, 84, 87), (45, 71, 76), (53, 94, 102), and (73, 103, 114) for the 10th, 50th, and 75th quantiles for four distributions. Historical data show that the true values of these quantiles are (63, 80, 89), (44, 70, 83), (52, 90, 110), and (72, 100, 115). From this information, the judgmental errors are computed as (9, 4, −2), (1, 1, −7), (1, 4, −8), and (1, 3, −1). It follows that the expert, on average, overestimates the 10th and the 50th quantile by 3 units each, but underestimates the 75th quantile by 4.5 units. Counter-adjusting for these biases, the random noise in the expert's estimates for the four calibration distributions are (6, 1, 2.5), (−2, −2, −2.5), (−2,1, −3.5), and (−2, 0, 3.5). We use these data to calculate the variance–covariance matrix of noise in judgmental errors as $\Omega = \begin{bmatrix} 16 & 2 & 5 \\ 2 & 2 & 1 \\ 5 & 1 & 12.33 \end{bmatrix}$. This matrix represents the reliability of the manager's judgments for various quantiles, for example, his judgments for the median have the least noise while his judgments for the 10th quantile tend to have the highest noise.

3.2. Step 2: Optimization-Based Model and Solution

The calibration information is then used to connect the expert's quantile judgments with the mean and standard deviation. We specifically seek to estimate the mean as the weighted linear average $\mu = w_1q_1 + w_2q_2 + w_3q_3$, and the standard deviation as a different weighted average $\mu = w'_1q_1 + w'_2q_2 + w'_3q_3$; where q_1, q_2, q_3 are the quantile judgments provided by the expert after deducting his biases, and w_i, w'_i; $i = 1, 2, 3$ are the weights to use on these judgments. This linear formulation is easy to use in practice and managers are familiar with this functional form as it closely resembles the formulas used for activity duration calculations in the PERT technique and in other applications (e.g., see Keefer and Bodily 1983). Our approach below determines these weights, given the manager's judgmental errors as captured in $\mathbf{\Omega}$.

We seek to obtain these weights such that (i) the variation in the estimated values of the mean and standard deviation is minimum, and (ii) on average the estimated values of the mean and standard deviation are equal to the true values. These two conditions are modeled in the form of a constrained convex minimization problem, and the solution of this problem is obtained as follows:

$$[w_1 w_2 w_3]^T = \boldsymbol{\Omega}^{-1}\mathbf{Z}(\mathbf{Z}^\mathbf{T}\boldsymbol{\Omega}^{-1}\mathbf{Z})^{-1}[\mathbf{1} \quad \mathbf{0}]^\mathbf{T},$$

$$[w'_1 w'_2 w'_3]^T = \boldsymbol{\Omega}^{-1}\mathbf{Z}(\mathbf{Z}^\mathbf{T}\boldsymbol{\Omega}^{-1}\mathbf{Z})^{-1}[\mathbf{0} \quad \mathbf{1}]^\mathbf{T},$$

where the matrix $\mathbf{Z} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{z_1} & \mathbf{z_2} & \mathbf{z_3} \end{bmatrix}^\mathbf{T}$, and each z_i corresponds to the standardized distribution. The z_i values are available for location-scale distributions, which includes the Normal distribution, the uniform distribution, Logistic, Pareto and Gumble distributions. For example, for the 10th, 50th, and 75$^{\text{th}}$ quantiles of the Normal distribution, these values are $z_1 = -1.28$, $z_2 = 0$, $z_3 = 0.67$. The weights obtained using the expressions above directly depend on the quality of the expert's judgments (captured by the calibration information in $\boldsymbol{\Omega}$) and the quantiles that the manager has chosen to provide. The calculations required to obtain the weights are easily performed in a spreadsheet environment. For the expert considered earlier the weights are obtained as $[w_1\ w_2\ w_3] = [0.03, 0.92, 0.05]$, and $[w'_1 w'_2 w'_3] = [-0.55, 0.12, 0.43]$.

3.3. Use of Weights for Distributions Without Prior Data

The analysis above provides the following protocol for using an expert's quantile-judgments to determine the mean and standard deviation of distributions that do not have prior data: First obtain the expert's judgments for the quantiles he selected and was calibrated on (e.g., the 10th, 50th, and 75th). Suppose these judgments are x_1, x_2, x_3. De-bias these judgments to obtain de-biased judgments q_1, q_2, q_3 by subtracting the bias (e.g., for the expert consider above, $q_1 = x_1 - 3$, $q_2 = x_2 - 3, q_3 = x_3 + 4.5$). Then use the weights determined above on the de-biased judgments to obtain the mean and standard deviation. These values then can be used as inputs to production planning and other operational models.

4. Benefits from Managerial Calibration: Evidence at Dow AgroSciences

The theory developed has been used at Dow AgroSciences for its annual production planning decision worth \$800 million for seed corn. The firm offers a few hundred varieties of seed corn in the market every year. Most of the seeds are new and have not been produced before, except for a few dozens that have consistently seen a high demand from farmers and have been produced repetitively in the last few years. The historical data for these seeds provide the calibration distributions. Each year an internal team at DAS calibrates the yield-expert's judgments for the 10th, 50th, and the 75th using the calibration distributions, and deduces the mean and standard deviation of the yield distributions for new seeds using the approach developed earlier. Then it uses the mean and standard deviation values as inputs to an optimization model that provides the optimal area to use to grow each variety of seed corn.

4.1. Monetary Benefits in Production Planning Context

The monetary benefit from using our approach stems from two sources, first, an estimate of the standard deviation is now available to make the tradeoff between using too much and too little production capacity, and second, the estimates of the mean and standard deviation explicitly incorporate the yield-expert's judgmental errors. To quantify the benefit on both accounts, we analyzed the production decisions made in 2014, using our approach. Analysis on these decisions showed that the annual production investment decreased by 6–7% over an earlier approach in use that ignored the uncertainty in yield and made the production decision assuming the yield-expert's judgment for the median to be the fixed yield. The profit also increased by 2–3%. These percentages amount to several millions of dollars on an annual basis for DAS. Equally important, DAS did not see a drop in the service levels of the seeds after the adoption of this new approach for estimating yield distributions. At the end of the annual selling season, the seed corn business manager commented that "this was the first time in many years that the crop plan required less capital investment than the amount allocated to DAS for this purpose, without a perceptible decrease in sales."

Subsequently, we repeated the analysis assuming that the yield-expert provided the quantile judgments but his judgments were not calibrated. Mathematically, this situation is represented by assuming that the biases in his quantile-judgments are equal to zero and the variance-covariance matrix Ω is an identity matrix. Our analysis showed that using the calibration information increases the expected profit by several millions of dollars. This analysis suggests that expert calibration leads to better decisions. Further the benefits from expert-calibration, when permitted by the availability of calibration distributions, can far exceed the effort required.

4.2. Objective Quantification of Manager's Expertise

The approach developed led to several non-monetary benefits as well. One question that frequently comes up during the use of expert judgments is whether it is possible to determine, "how good are the manager's judgments" into a metric. The approach developed provides an objective metric for this evaluation. Specifically, our approach provides the estimates of the mean and standard deviation using an expert's judgments; with some additional calculations it also provides the confidence interval around these estimates. Using results from statistics, we can determine the size of a random sample of data that would have provided estimates of the mean and standard deviation with the same width of confidence intervals. The expert (or equivalently his judgments) is deemed to be equivalent to this sample size. At DAS we determined that the yield-expert's quantile judgments are equivalent to 22 data points. This size is equivalent to approximately five to six years of test data at DAS, and exceeds the data points obtained during the complete life-cycle of most seeds.

5. Summary of Contributions

We develop an approach to deduce the mean and standard deviation of probability distributions using expert judgments for quantiles. This approach is analytically tractable and is easy to use in practice. It provides the flexibility of using the judgments for any set of quantiles that an expert is willing to provide. The approach also establishes a novel equivalence between the quality of an expert's judgments and the size of an experimental sample that is equally informative about the distribution. Finally, we note that the approach developed is extendible to multiple experts and Bayesian updating in a straight forward fashion. An extended set of technical details of the approach described in this article are available in Bansal et al. (2016).

References

Bansal, S., G. J. Gutierrez, J. R. Keiser. 2016. Quantifying uncertainties using expert assessments in a dynamic new product development environment. Working paper, The Pennsylvania State University.

Henig, M., Y. Gerchak. 1990. The structure of periodic review policies in the presence of random yield. *Oper. Res.* **38**(4): 634–643.

Keefer, D. L., S. E. Bodily. 1983. Three-point approximations for continuous random variables. *Management Sci.* **29**(5): 595–609.

O'Hagan, A., J. E. Oakley. 2004. Probability is perfect, but we can't elicit it perfectly. *Reliab. Eng. Syst. Saf.* **85**(1–3): 239–248.