Innovative Applications of O.R.

What to do when decision-makers deviate from model recommendations? Empirical evidence from hydropower industry

Anssi Käki[a], Katariina Kemppainen[b,*], Juuso Liesiö[b]

[a] *UPM-Kymmene Corporation, P.O. Box 380, 00101 Helsinki, Finland*
[b] *Department of Information and Service Management, School of Business, Aalto University, P.O. Box 21210, 00076 Aalto, Finland*

ABSTRACT

Decision makers do not always follow recommendations from model-based decision support systems. We suggest that analyzing the differences between decision recommendations produced by prescriptive models and the behavior of the decision makers provides valuable insights that can be utilized to improve model-based decision support processes. Specifically, we develop an intervention process in the context of hydropower production planning to study the motivations of decision makers and the ramifications of their behavior. The analysis is based on deviations between recommendations of an in-house optimization tool and actual decisions, enhanced by planner feedback collected from a daily web-survey. We find that even though the planners make some adjustments with positive financial impact, their actions mainly worsen the performance of the production plan. Using the collected data, we identify several reasons for the deviations and recommend multiple enhancements to the planning process. For example, we propose a shift from output-adjusting to input-adjusting interaction between human planner and model. Altogether our facilitated modeling project shows that combining objective and judgmental process feedback is superior for recognizing corrective actions and systematically improving model-driven decision processes. Furthermore, the intervention process developed for this case gives structure for the lifecycle management of model-based decision support systems.

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

With the rise of analytics and increasing number of data sources, decision support models are seen as the prominent source of competitiveness especially in complex and volatile business environments (Vidgen, Shaw, & Grant, 2017). It follows that practitioners are increasingly interested in gaining insights from data with OR tools such as mathematical optimization, simulation and statistical analysis (Mortenson, Doherty, & Robinson, 2015). The expectations are building up also due to various reported success stories in the OR academic competitions by EURO (Excellence in Practice Award), INFORMS (Franz Edelman Award), and POMS (Martin K. Starr Award). As a result of this data-driven analytics movement, companies are increasingly exposed to the challenge of designing, implementing and managing decision making processes to maximize the business value produced by these new models (Liberatore & Luo, 2010).

It is commonly agreed that the use of analytics can benefit decision making by economizing on cognitive effort, solving cognitively intractable problems, producing insights, integrating knowledge, and managing conflict (see, e.g., Liberatore, Hatchuel, Weil, & Stylianou, 2000; Luoma, 2016). However, many authors have suggested that decision-makers (DMs) systematically deviate from recommendations produced by decision models. For instance, Stoop and Wiers (1996) report experiences from scheduling, where the planners believe that by using enough mental effort they can obtain plans that outperform those produced by the imperfect planning system. Van Donselaar, Gaur, Van Woensel, Broekmeulen, and Fransoo (2010) observed that retail store managers consistently modified the order recommendation generated by the automated replenishment system by advancing orders from peak days to non-peak days. Elmaghraby, Jank, Zhang, and Karaesmen (2015) show that B2B pricing decisions do not automatically correspond to the recommendations of a pricing tool. Petropoulos, Fildes, and Goodwin (2016) report that experts' adjustments to model-generated forecasts resulted in a 14% drop in forecasting accuracy.

Identifying reasons for deviations can be very valuable for process improvement, as the existence of deviations suggests that

* Corresponding author.
 E-mail addresses: anssi.kaki@upm.com (A. Käki), katariina.kemppainen@aalto.fi (K. Kemppainen), juuso.liesio@aalto.fi (J. Liesiö).

decision making processes do not perform optimally or according to DM's expectations. In particular, if model recommendations are consistently better than actual decisions, more rigorous use of model recommendations would improve the process outcome (e.g., Bolton & Katok, 2008). If, in turn, the DM systematically outperforms model recommendations, then the model should be improved to automate the DM intervention thus freeing DM's cognitive resources for alternative uses (Kayande, De Bruyn, Lilien, Rangaswamy, & Van Bruggen, 2009; Luoma, 2016). Consequently, this paper studies how analyzing deviations between model recommendations and actual decisions taken by DMs can be utilized to improve both decision support models and decision making processes. More specifically, we report experiences from an intervention deployed in the production planning process of a European hydropower producer. Recently, the company found out that even though their in-house developed planning system was used frequently, the actual production plans deviated from the optimized plans significantly. Thus, we set out to analyze daily planning decisions – each of which considers electricity worth 10,000–100,000 euros sold to a spot market. The daily data set includes (i) model-generated decision recommendation, (ii) actual decision taken by planners, (iii) planner's subjective reasoning/justification for deviations, and (iv) ex-post observations of key uncertainties. This data allows us to evaluate and compare the performance of both human (DM) and decision support system (DSS), i.e., model recommendations with and without planner's intervention.

The results from this case study suggest that even though the planners make some adjustments with positive financial impact, their actions mainly worsen the performance of the production plan. But there are multiple reasons for this behavior, some of which are out of planners' control. For example, on one day, the planner disagrees with a price forecast and makes a risky deviation to increase revenues, whereas on another day, the manager instructs the planner to mitigate risk of violating the environmental constraints. By identifying the causes of deviations, our intervention process helped the company identify actions to improve the performances of their hydropower planning process. These actions include extensions to the optimization model and improving the quality of its input data as well as changes to the manner in which the human planners interact with the in-house model.

We anticipate that the intervention process developed for this case can be used in other context as well to improve model-based decision making processes. In the spirit of facilitated modeling (Franco & Montibeller, 2010) this process involved key managers in the client organization and included the collection of both objective decision process data (e.g., model input/outputs, actual decisions) and judgmental feedback from the DMs (planners). Instead of analyzing only one or the other, combining objective and judgmental data arguably provides richer and more accurate understanding of how decision processes supported by DSS work and how those can be improved. Analyzing objective data alone can be misleading: for example, it might indicate that a planner refuses to follow model recommendations when, in fact, he would like to follow them but organizational issues prevent it (manager's instructions; unwritten rules among peers). Then again, interviewing decision makers without looking at objective data is also short-sighted: for example, in forecast-driven decisions planners can justify their actions by expressing distrust at the forecast, but hindsight analysis might reveal that following the forecast would have yielded superior results compared to realized ones. In general, tracking process performance and deviations by DMs is a critical piece in improving a model-based decision processes. Moreover, feedback-driven improvement should, in the long run, help ensure that model-driven DSS generates value.

This study also intersects two other streams of research worth acknowledging. First, the paper contributes to the growing liter-ature of Behavioral Operational Research (BOR) that investigates behavioral issues in the use of OR models (Franco & Hämäläi-nen, 2016; Hämäläinen, Luoma, & Saarinen, 2013). Specifically, our study sheds light into the role of OR models in decision making and how decision makers interact with those models. Importantly, unlike controlled laboratory experiments, our intervention focuses on a real business decision process, and thus is imbedded in real organizational context. The importance of organizational realities in how organizations adopt and use OR methods has been recently emphasized for example by Luoma (2016). Luoma introduces a taxonomy that can be used to identify organization-specific factors that have contributed to the success or failure of specific modeling activities. In the taxonomy, the decision process of this study falls under "routine-decision making" as opposed to "problem solving", because hydropower production decisions are repeated each day.

Second, this paper can be categorized as a study in Behavioral Analytics, which Durbach and Montibeller (2019) characterize as the use of behavioral data sets storing judgements and choices to advance BOR. In particular, behavioral analytics can aim at (i) detecting behavioral patterns, (ii) exploiting behavioral findings, and (iii) improving judgement. The intervention reported in this paper addresses all these three objectives. Arguably, the main driver of behavioral analytics is the increasing number of websites that record users' judgement and choices (e.g., online shopping and games) which are open to the general public. However, our study suggests that additional opportunities for Behavioral Analytics are offered by the increased use of analytics to support decision making in businesses, which enables organizations to generate internal databases storing model recommendations and actual decisions made.

The rest of this paper is organized as follows. Section 2 introduces literature for studying and improving model-based decision processes. Section 3 reports how we studied the performance of production planning process of a hydropower producer. Section 4 reports the main results and recommended actions, and Section 5 discusses managerial implications in general. Section 6 concludes by discussing the limitations of our approach and by presenting future research topics.

2. Research on improving model-based decision processes

In general, there are two modes for human-model interaction, as illustrated in Fig. 1, but these are not mutually exclusive. The first, *input-adjusting interaction*, represents the conventional OR approach where the DM adjusts model parameters (or inputs) to explicitly consider, for instance, particular decision circumstances or risk attitudes. The second human-model interaction type is *output-adjusting* where the model output is adjusted directly by the DM. Arguably input-adjusting is the correct way to use models for decision support but in reality, output-adjusting is common. For example, Fransoo and Wiers (2006) report experiences from a Material Resources Planning system, where planners frequently modify orders in the plan generated by a DSS even though the system offered the possibility to change the parameters. Based on experiences from infrastructure management, Mild, Liesiö, and Salo (2015) suggest that it can indeed be easier for the DM to utilize tacit or holistic knowledge on decision alternatives by modifying the model recommendation directly.

Studying the performance of a decision process empirically is more straightforward if decision process follows input-adjusting approach, i.e. the actions of DM are recorded in input data changes and thus, objective data can be sufficient to explain the DM behavior. In a repeated decision setting, such as the one studied in this paper, data sets containing detailed records of the changes the DM has made to the model inputs could be directly compared to the
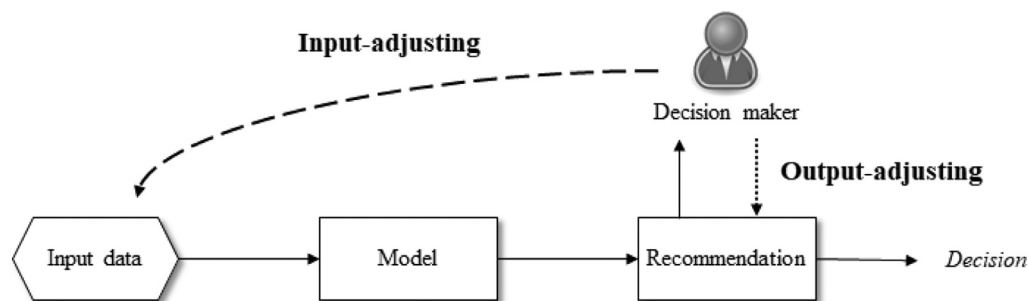
Fig. 1. Two types of human-model interaction.

ex-post observed actual values to evaluate whether these changes increase or reduce errors.

Data sets from repeated input-adjusting decision processes can be used to analyze a large number of well-known biases demonstrating how human judgement and choices frequently deviate from normative rules. Research on these biases can be traced back to the seminal paper of Tversky and Kahneman (1974). Montibeller and von Winterfeldt (2015) provide a recent summary of this literature and develop an exhaustive list of both cognitive and motivational biases that can affect decision support and risk analysis. Cognitive biases result from the heuristics and shortcuts humans utilize in their mental processes: for instance, plenty of empirical evidence shows that results from estimation of numerical values are affected by given initial values (anchoring bias). Motivational biases, which can be both conscious and subconscious, result from self-interest, social pressures or the organizational context. For instance, empirical evidence shows that the probabilities of favorable outcomes are often overestimated (optimism bias). Montibeller & von Winterfeldt also recommend techniques for reducing or removing each type of bias.

If model's outputs are adjusted instead of its inputs, available objective data only reveals dissatisfaction in decision recommendation and shows the corrective action, not the reasons behind them. Hence, tracing the source of these adjustments and linking it to a particular bias can be extremely difficult. Here, a typical scientific approach is to form a hypothesis of the behavior and test it with decision process data. For example, Van Donselaar et al. (2010) analyzed managers' order behavior in retail stores under the hypothesis that managers consistently modified the order recommendation (of the automated replenishment system) by advancing orders from sales peak days to non-peak days. They found support for this and were able to explain the DM behavior with misaligned incentives: the automated system minimized inventory holding costs, while the managers focused on minimizing in-store handling costs and stock-outs. Later, Elmaghraby et al. (2015) described how salespeople's pricing decisions are affected by a price recommendation system by developing hypotheses for relationships between recommended and observed prices. They found that in some cases, the recommendation system impacted more on the decision to change price (from previous negotiated price) than the absolute value of change. In both papers above, the authors were able to find explanations for human behavior in output-adjusting processes by using objective data alone. However, in many business environments this is very difficult as hypothesis development is best applied to rather simple decision settings.

Also controlled laboratory experiments are increasingly used to better understand decision making processes. Newsvendor experiments study a stylized setting for ordering under demand uncertainty where decisions are made without any model support in a laboratory setting: this literature stream started with Schweitzer and Cachon (2000), followed by e.g. Bolton and Katok (2008) who study the impact of experience and feedback on Newsvendor deci-

sions, and Gavirneri and Isen (2010) who offered subjects the possibility to ask for more information during the experiment. Käki, Liesiö, Salo, and Talluri (2015) find indicative evidence that experts (consultants) perform better than students in newsvendor setting because they used spreadsheets based tools to support decision making. Hoch and Schkade (1996) study the use of DSS for forecasting and find in their experiment that use of models, when used in combination with expert knowledge, provide superior and robust decision support. Kayande et al. (2009) study a setting where subjects make complex marketing decisions and find that aligning the decision maker's mental model with the DSS's underlying model is critical for decision performance. They emphasize the importance of feedback in closing the "acceptance gap" between the DM and the DSS. All abovementioned experiments offer rich perspective to DM behavior, but leave open the question how well the results carry into more complex (and realistic) decision settings, as in our case.

Fransoo et al. (2010) cover a wide array of behavioral aspects related to planning and scheduling in a complex manufacturing setting. Wäfler et al. (2010) describe different frameworks for understanding control behavior of a planner using a DSS for manufacturing. They emphasize the importance of empowering the decision maker to really perform control: this is why in a typical planning process the DM has a mandate to adjust the model recommendation as she sees fit (as is the case with our case company, too). Riedel et al. (2010) discuss the importance of user acceptance for the quality and results of DSS-driven planning process. They posit that acceptance and trust can (and should) be taken into account in DSS building phase, leading into a better overall decision quality. Riezebos et al. (2010) emphasize that designing scheduling algorithms should not be based purely on the development of a scientific model for the scheduling problem at hand, but largely on human and organizational characteristics as well.

To summarize, our intervention process is based on collecting and analyzing objective data about model recommendations and actual decisions from a real-life decision process similar to Van Donselaar et al. (2010) and Elmaghraby et al. (2015). However, our process systematically utilizes judgmental data provided by planners per each decision in order to gain insights on the DM behavior. This study also complements the field study book edited by Fransoo et al. (2010): our process is developed for improving existing model-based decision processes, whereas those field studies mostly focus on defining the design principles of new DSSs.

3. Empirical study in hydropower production planning

Our study focuses on a river system with four hydropower plants that produce electricity worth some $50,000 each day (with variation from $10,000 to over $100,000). The company operating the hydropower plants has developed an in-house optimization model for production planning but the planners are not obligated to follow the model's recommended plan. Indeed,

deviations from the optimized plan happen regularly. The company managers wanted to understand the motivation behind these deviations, whether these deviations are beneficial or harmful, and if the company should seek to reduce the deviations or learn from them. Against this backdrop, the analyst team (consisting of the authors, one of them working in the company) specified the following research objectives:

1. Identify key drivers behind the planners' decisions to adjust the optimized plan,
2. Evaluate the financial impact of the deviations, and
3. Recommend how to improve the planning process.

In order to pursue these objectives three areas were identified for detailed analysis: the planning process (business environment, organization, schedule), the optimization model (inputs, outputs, model logic, user interface), and the operational data available from the process. Before the data collection and analysis, some half-a-dozen planners and managers were interviewed (unstructured; 1–2 hours each). The interviews and existing data indicated the existence of deviations but were not sufficient to track the underlying root causes. A particular data gap was identified: reliable analysis of planners' motivations for deviating from the optimized plans was not possible with the existing data or using the interview insights. Therefore, a tailored web-survey was developed to collect planner feedback daily during the study period. In the following sections, we describe these three areas (process, model, data) in more detail.

3.1. Planning process

Hydropower planning offers an attractive area to study model-based decision processes. In particular, it is a repeated decision making setting, where some aspects of decisions made can be evaluated ex-post based on observed quantitative measurement, such as electricity prices. Moreover, it is a mature field from the modeling perspective with an extensive literature on how optimization can be used to identify better plans (e.g., Fleten & Kristoffersen, 2007; Fosso, Gjelsvik, Haugstad, Mo, & Wangensteen, 1999; Wallace & Fleten, 2003). Yet, regardless of the detailed models available, fully automated planning solutions are rare because of the related complexities and responsibility considerations and thus, hydropower producers widely employ planners (cf. decision makers) to manage this operation. Hence, behavioral aspects play a noteworthy role in hydropower planning but the decision process has not been widely studied before.

The company sells electricity to a spot market, in which next-day physical delivery prices and volumes are determined through a double auction for each of the 24 hours (for a general description of such market, see Flatabø, Doorman, Grande, Randen, and Wangensteen, 2003). Company's next day production plan is optimized against a price forecast and offered to the market with a price that guarantees bid acceptance. In practice, this translates to a capacity constrained production planning process under price uncertainty and unlimited demand. The short-term planning process is repeated every day and the binding decision horizon covers 24 hours; the full short-term planning horizon is dynamic and covers at least few days and up to a week to account for the future value of production. The end condition for the short-term model comes from a 12-month mid-term planning process of hydro reservoirs, which is not in the scope of this paper. In the short-term planning process under study, the planner can utilize an optimization-based decision support system.

The planning process is run by three planners, one of whom at a time is responsible for next day's production plans. The planners are engineers with academic degrees but without formal training in operations research. The process (see Fig. 2) starts around 8AM

Table 1
Model notation.

Indexes	
$i = 1 \ldots N$	Set of power plants
$t = 1 \ldots T$	Set of hours in the planning horizon
Decision variables	
$q_i(t)$	Discharge at plant i at hour t
$x_i(t)$	Storage volume at reservoir of plant i at hour t
Note storage volume can be	
x_i^{OVER}	Storage above target at reservoir of plant i at final hour T
x_i^{UNDER}	Storage below target at reservoir of plant i at final hour T
Parameters	
$I_i(t)$	Forecasted inflow to reservoir at plant i at hour t
$\hat{p}(t)$	Forecast for electricity price at hour t
ϕ_i	Energy generation function for plant i
$\alpha_{i,d}$	Delay factor for upstream discharge d hours ago to reservoir of plant i, which satisfy
$\sum_{i,d} \alpha_{i,d} = 1, \ \alpha_{i,d} \geq 0.$	
C_i	Penalty of over/under storage in reservoir of plant i
X_i^{TARG}	Target level for reservoir of plant i at the end of planning horizon
$\underline{Q}_i(t), \bar{Q}_i(t)$	Minimum and maximum discharge at plant i at hour t
$\underline{X}_i(t), \bar{X}_i(t)$	Minimum and maximum storage level at plant i at hour t

from the initialization of the optimization model. It entails fetching price and inflow forecasts, plant maintenance schedules, reservoir levels, and other critical data from the energy management system. The optimization is then run, typically only once, after which the planners can adjust the plan prior to generating a spot sales offer due at 10:00AM. After this, the planner moves on to plan other river systems in the portfolio. In the afternoon, after prices realize, there are couple of hours for analysing and planning next day's operations. Overall, the process is subject to a strict timetable that causes time pressure for the planners.

3.2. Optimization model

The company has developed the optimization model over the past few years to maximize profit. The planning period can range from a few days up to ten, depending on the weekday. The model ensures plan feasibility by taking into account various technical aspects, such as environmental limits for reservoir levels and water flow rates, water travelling times between power plants and the energy generation function. Summary of the model notation is presented in Table 1.

Specifically, the model seeks to identify hourly discharge levels for each plant that maximize the total revenue based on forecasted electricity prices. For plant $i \in \{1, \ldots, N\}$ the decision variables $q_i(t)$ and $x_i(t)$ capture the discharge and reservoir storage volume at hour $t \in \{1, \ldots, T\}$, respectively. These variables are linked through the hydro balance constraint

$$x_i(t) - x_i(t-1) = I_i(t) + \sum_d \left[\alpha_{i,d} \ q_{i-1}(t-d) \right] - q_i(t) \quad \forall i, \forall t,$$

(1)

which ensures the change in the reservoir volume $x_i(t) - x_i(t-1)$ is equal to the difference between the plant's discharge $q_i(t)$, and incoming flows from the environment I_i as well as the those discharged by upstream plant indexed $i-1$. This discharge reaches plant i with a delay which is captured by the parameters $\alpha_{i,d}$. For instance, parameter values $\alpha_{2,1} = 0.1$, $\alpha_{2,2} = 0.5$, and $\alpha_{2,3} = 0.4$ imply that the flow to plant $i = 2$ at hour t contains 10% of the discharge plant $i-1 = 1$ produced at hour $t-1$, 50% of discharge produced at hour $t-2$, and 40% of discharge produced at hour $t-3$.
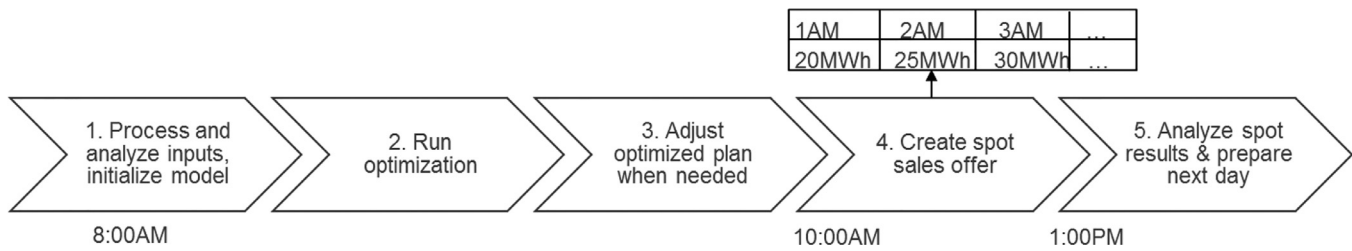
Fig. 2. Schematic planning process for daily hydropower production decisions.

Existing environmental permits imply interval bounds for the hourly discharge and storage-volume levels, which are denoted by $[\underline{Q}_i(t), \bar{Q}_i(t)]$ and $[\underline{X}_i(t), \bar{X}_i(t)]$, respectively. Moreover, the company's mid-term planning process sets target levels for the reservoir volumes X_i^{TARG} that should be met at the end on the 7-day modeling period ($t = T$). These considerations lead to additional linear constraints

$$\underline{X}_i(t) \leq x_i(t) \leq \bar{X}_i(t) \quad \forall i, \forall t \tag{2}$$

$$\underline{Q}_i(t) \leq q_i(t) \leq \bar{Q}_i(t) \quad \forall i, \forall t \tag{3}$$

$$x_i(T) + x_i^{UNDER} - x_i^{OVER} = X_i^{TARG} \quad \forall i \tag{4}$$

$$x_i^{UNDER}, x_i^{OVER} \geq 0 \quad \forall i, \tag{5}$$

where the auxiliary decision variables x_i^{UNDER} and x_i^{OVER} capture deviations from the storage target level X_i^{TARG}.

The energy produced at each plant depends on both the discharge $q_i(t)$ and the storage-volume $x_i(t)$: higher discharge increases production because of larger masses rotating the turbine, and produced energy is higher on high storage levels because this implies higher hydraulic head and thus more potential energy. Due to typical hydro turbine characteristics, the energy production suffers from diminishing returns with respect to discharge, meaning that the higher the discharge, the less efficient production becomes. In the optimization model the energy generation function for each plant i is approximated with a concave piecewise linear function $\phi_i(q_i(t), x_i(t))$, which can be formulated using linear constraints and continuous auxiliary decision variables with standard techniques (for details see, e.g., Nemhauser & Wolsey, 1988, pp. 11).

The objective function – in addition to forecasted revenue from energy sold to the market – includes a term penalizes from deviating from the storage target levels. Hence, the optimization model can be formulated as:

$$\max_{q,x} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \hat{p}(t) \phi_i(q_i(t), x_i(t)) \right) - C_i \left(x_i^{UNDER} + x_i^{OVER} \right) \right]$$

subject to constraints (1)–(5), where parameter $\hat{p}(t)$ is the electricity price forecast for hour t and parameter C_i captures the penalty cost for deviation from the storage targets at plant i.

Some technical details are omitted from this formulation without losing the essential features of the problem. For example, (i) the model allows bypassing the turbines by spilling in flood-like situations, (ii) one out of the four plants operates in an automated mode, and its discharge is a function of an upstream plant's discharge, (iii) the model has optional penalty for too high hour-by-hour changes in storage volumes, and (iv) it is possible to account for additional revenue sources such as the balancing markets. Altogether, the model in use includes some ten thousand decision variables, most importantly the discharges $q_i(t)$, the storage volumes $x_i(t)$ and the auxiliary variables required to formulate the

piecewise-linear energy production functions $\phi_i(\cdot, \cdot)$. However, the model can be solved with standard Linear Programming (LP) algorithms in few seconds using a standard laptop. The implementation is done with C#, and the end-user interface is in Microsoft Excel.

3.3. Data collection

The company has a process to collect and store detailed data related to the optimization model (cf. objective data). We analyzed data from 290 days between November 2015 and August 2016. However, in total 61 days were excluded from further analysis due to three flood-like hydrological periods during which all plants had to be operated at practically maximum power, and hence planning process did not involve any real decision making. From the remaining 229 days, we had access to hourly optimization inputs, optimization results, and actual decisions. During the analysis period, the planners used the mandate to adjust optimized plans regularly as exemplified by Fig. 3, which shows optimized and actual plans for a 24-hour period.

Planner feedback (cf. judgmental data) was collected daily via web-based survey. The survey content was based on interviews of company planners and managers. The interview process was semi-structured and discussion evolved around three topics: (i) the current planning process and tool, (ii) reasons for not using tool's results directly, and (iii) improvement ideas related to the process. The interviewees included two planners, their manager and one developer. They all had varying assumptions about why deviations from the optimized plans took place. For example, one experienced planner stated that the price forecast was not always aligned with his own views of the market, which led him to make occasional changes. Another planner emphasized that the plans often had to be adjusted for external reasons, such as plant maintenance. The manager who had led the development of the optimization tool, in turn, emphasized that the model did not account for, e.g., the fixed cost of a generator start-up, which is likely to cause planners to minimize start-ups through their own adjustments.

Based on the interview insights, a two-part survey was created. The compulsory part consisted of three Likert-scale assessments: (i) general market uncertainty (1 = not uncertain, 5 = very uncertain), (ii) how busy the planner was (1 = very busy, 5 = not busy at all) and (iii) how satisfied the planner was with his decision (1 = very dissatisfied, 5 = very satisfied). It also had an optional part to be filled in case the planner deviated from the optimized plan. In these cases, the planner answered questions of type "Did factor X have an impact on your decision to change the plan?" There were altogether nine alternative factors (X) for deviations under three aggregate categories:

i. **"Market"**: Distrust in the price forecast's hourly profile; Distrust in the price forecast's daily profile; Preparing for intra-day market actions; Preparing for other markets

ii. **"Technical"**: Reducing hour-by-hour changes in production plan; Reducing the risk of too high/low reservoir levels; Reducing the number of required generator start-ups
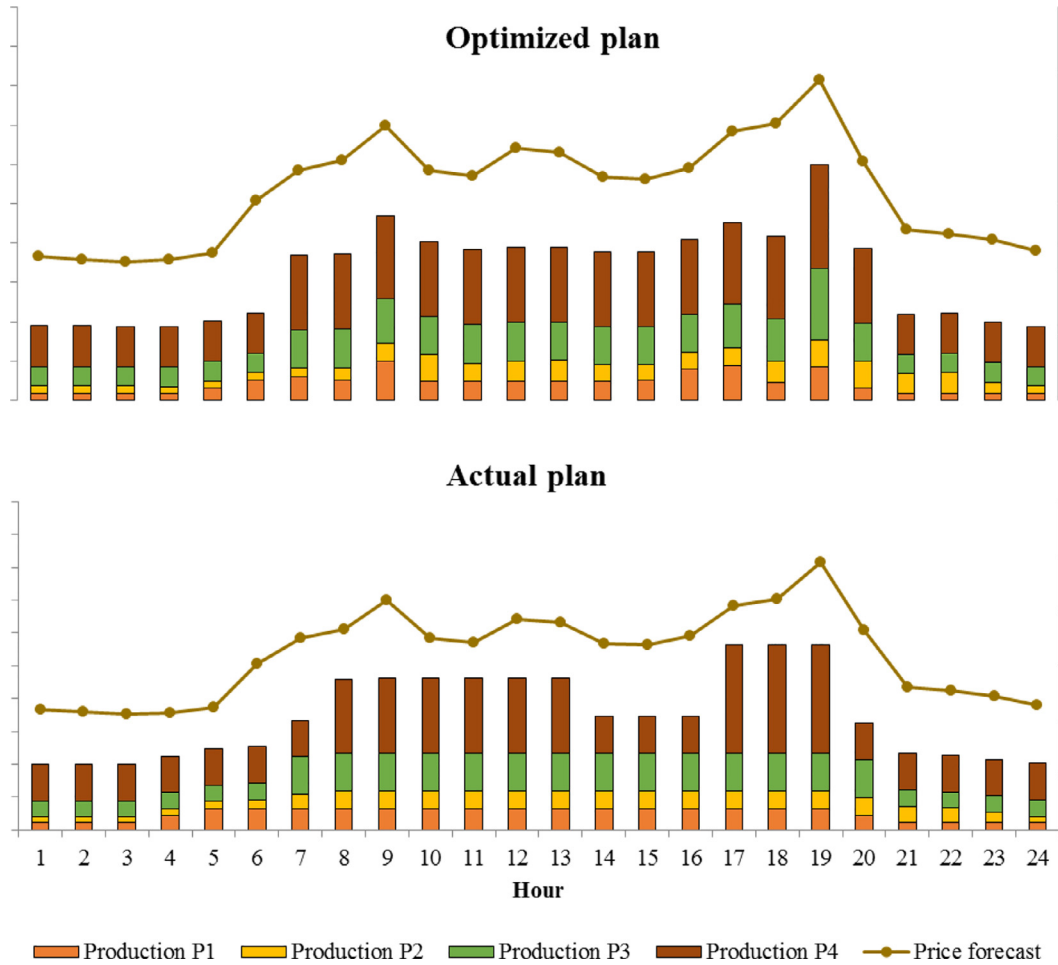
Fig. 3. Optimized and planner's actual plan for four power plants (P1–P4) during a 24-hour-period.

iii. **"External"**: Unscheduled and planned maintenance at plants; Manager's instructions

All of these questions were answered on a Likert five-point scale (1 = no impact, 5 = significant impact). The planners could also leave any of their answers empty. Based on discussions with the planners we decided that an empty answer corresponds to "1 = No impact".

The on-line survey was filled on 144 (63%) days. The missing feedback is mostly due to forgetting to fill in the survey, or being too busy with other work tasks. Filling the survey form was not incentivized, which partially explains the response rate. Whenever a planner responded 4 or 5 to any of the factors in one category, we have marked the aggregated category as significant in explaining planner's actions.

3.4. Results

Next, we report our analysis based on all the objective and judgmental data we collected during the process. For more details, see Table 2.

3.4.1. Survey results

The survey responses are summarized in Table 3. They reveal that the planners feel some but not excessive time pressure (3.5/5) and that they are somewhat satisfied with their plans (3.5/5). Market uncertainty, though not particularly high overall, was still 30% of the time evaluated "high" (4) or "very high" (5). In 9% of days when the form was filled, the planners deviated from the optimal

plan due to a "Market" factor. Other factors, i.e., "Technical" or "External", were identified more often. None of the three factors was identified on 98 days, i.e., on those days the planner was not able or did not want to share why he deviated from the optimum.

3.4.2. Summary of deviations between optimized and actual plan

To quantify the differences between the optimal plan ($\phi^O(t)$) and the actual production plan ($\phi^A(t)$) of all four plants on a daily (24-hour) level, we use the mean absolute percentage error (MAPE) and mean percentage error (MPE):

$$MAPE = \frac{1}{24} \cdot \sum_{t=1}^{24} \frac{|\phi^P(t) - \phi^A(t)|}{\phi^P(t)}$$

and

$$MPE = \frac{1}{24} \cdot \sum_{t=1}^{24} \frac{\phi^P(t) - \phi^A(t)}{\phi^P(t)}$$

The average MAPE across all 229 days is 17%, meaning that on an average day, hourly productions quantities differ by, in average, 17%. The distribution of MPE between hours of the day in Fig. 4 shows that especially during typical night hours (23–07), planners systematically increase production whereas during day time, production is decreased from the optimized plan. Several forces are at play here.

First, during night time when prices systematically are the lowest the optimization model recommends a plan with lowest possible production. In reality, it is not always technically possible to

Table 2
Overview of the data used in the analysis.

	Description	Resolution	Number of obs.
Optimized plan	Outcome of the optimization model	Hourly	5496 hours (229 days)
Actual plan	Planner's decision	Hourly	5496 hours (229 days)
Price forecast	Forecast that was used in optimization and planner's decision making	Hourly	5496 hours (229 days)
Realized price	Market price that realized	Hourly	5496 hours (229 days)
Compulsory feedback form data	Assessment of market uncertainty; how busy planner was; how satisfied planner was	Daily	144 days
Optional feedback form data	True, if planner gave a 4 or 5 for at least one question in a category ("Market", "Technical", "External"), otherwise False	Daily	14 to 68 days depending on category
Textual feedback	Free form description of planning	Daily	107 days

Table 3
The mean and the share of 4's and 5's is calculated from 144 days when a form was filled, and the answer-% refers to the number of times any answer was given in a category.

	General (compulsory)			Reason for deviation (optional)		
	Is market uncertainty high?	Did you have enough time for planning?	Are you satisfied with your plan?	"Market"	"Technical"	"External"
Mean	3.2	3.5	3.5	1.4	2.6	3.1
%-of 4's and 5's	30%	48%	64%	9%	41%	47%
Answer-%	100%	100%	100%	47%	77%	70%
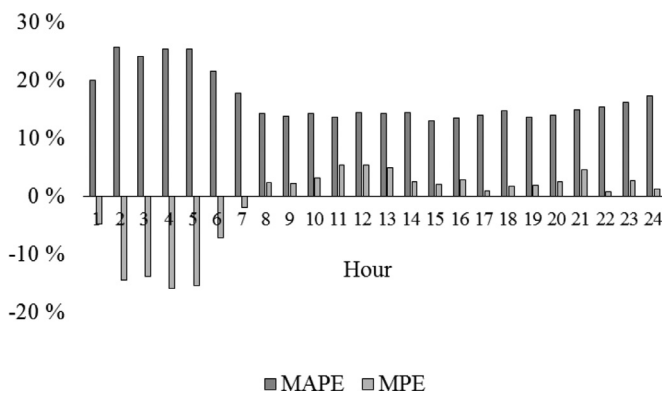
Fig. 4. Deviations between optimized and actual plans as averages for each hour over the 229-day data set.

drop production to a minimum level, and hence the planners have to adjust upward (on average 20% for hours from 1AM to 5AM).

Second, the planners may also want to avoid running the plants at maximum power under high production quantities because hydropower turbines have bad utilization rates in maximum levels. The optimization model, however, does account for these efficiency losses (by deploying the piece-wise linear energy generation functions ϕ_i), and hence the recommendation for maximum output implies that the price forecast is high enough to justify this action.

Third, planners also prefer "smooth" plans compared to the recommendations of the optimization model, which can make quite "spiky" plans due to relatively small hour-by-hour differences in the price forecast. In these cases, the planner can do "peak shaving", i.e., decrease production spikes.

3.4.3. Financial performance between optimized and actual plan

Over the period of 229 days, the planners' actual decisions amount to roughly 0.8% less sales than what the optimized plan would have achieved which translates to hundreds of thousands EUR. This result (as all other financial figures here) is based on actual prices, not the forecast $\hat{p}(t)$. The corresponding impact on sales using forecasted prices is around 1.0%, which means that the economic performance of optimized vs. planner's plan is not sensitive to forecast errors. In other words, potential forecast error does not favor one or the other in large scale. To analyze further how the financial performance varies day-by-day, we use the achieved

unit price (daily sales divided by daily production, [€/megawatt hour]). Specifically, we use the daily average difference between the optimization's achieved price (p^O) and planner's achieved price (p^A) as a single measure for the performance difference. This is denoted by $\Delta p = p^A - p^O$. Using achieved prices instead of, e.g., sales allows comparison between days where total volumes (and thus, sales) vary a lot from each other.

The average Δp over all 229 days is 0.41, which means that, on an average day, the optimization-based plan yields 41 cents higher price for each MWh (megawatt hour) sold. Yet, as can be seen in Fig. 5, the average Δp changes significantly when different filters are applied. On days when the on-line form was filled, planners' adjusted plan was significantly worse than optimization compared to the case, when form was not filled ($\Delta p = 0.64$ vs. $\Delta p = 0.19$). This is presumably a result of planner's eagerness to explain his actions when he thought that the adjustment was expected to be particularly harmful in economic sense. This is most obvious in cases when an "External" factor was reported as the reason for adjustment: here, $\Delta p = 0.92$, which is a large and statistically significant difference compared to rest of days for which, in this case, $\Delta p = 0.30$. On days when "Technical" factors are indicated as the reason, the average $\Delta p = 0.58$ meaning that the optimized plan would have given significantly better results, but this difference to the baseline ($\Delta p = 0.45$) is not statistically significant. Interestingly, deviations due to "Market" factors results in negative $\Delta p = -0.19$ which indicates that, in these days, the planner has outperformed optimization. The difference to baseline $\Delta p = 0.51$ is statistically significant.

We note that these results comply with a multiple regression model that was fitted to aggregated survey results to explain Δp. The amount of data was not sufficient to fit interaction terms, but in simple multiple regression, coefficient values and significances were in line with pairwise comparison discussed above and illustrated below (Fig. 5).

In summary, the planners' deviations can be classified to four cases: (i) deviations due to "Market" based reason; rare events when the planners outperform optimization, (ii) deviations due to a "Technical" reason that do not significantly worsen the planner's performance from the baseline but in which optimization still outperforms planners, (iii) deviations due to "External" issues, which significantly worsen the planners' performance, and (iv) cases where no reason was identified and in which the planners perform, on average, worse than the optimized plan ($\Delta p = 0.19$) but not as clearly as in cases in which a reason was identified.
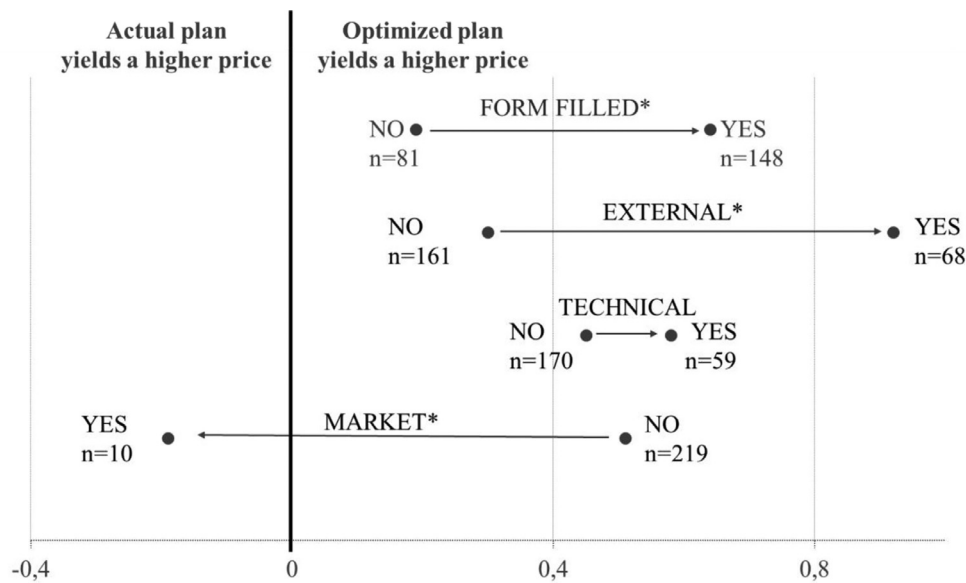
Fig. 5. Comparison of the financial performance of the actual vs. optimized plan, measured with $\triangle p$ in four cases: (i) when the web survey was/was not filled, or when (ii) external, (iii) technical or (iv) market based reason for deviation was (not) given. The asterisk indicates statistical significance ($p < 0.05$).
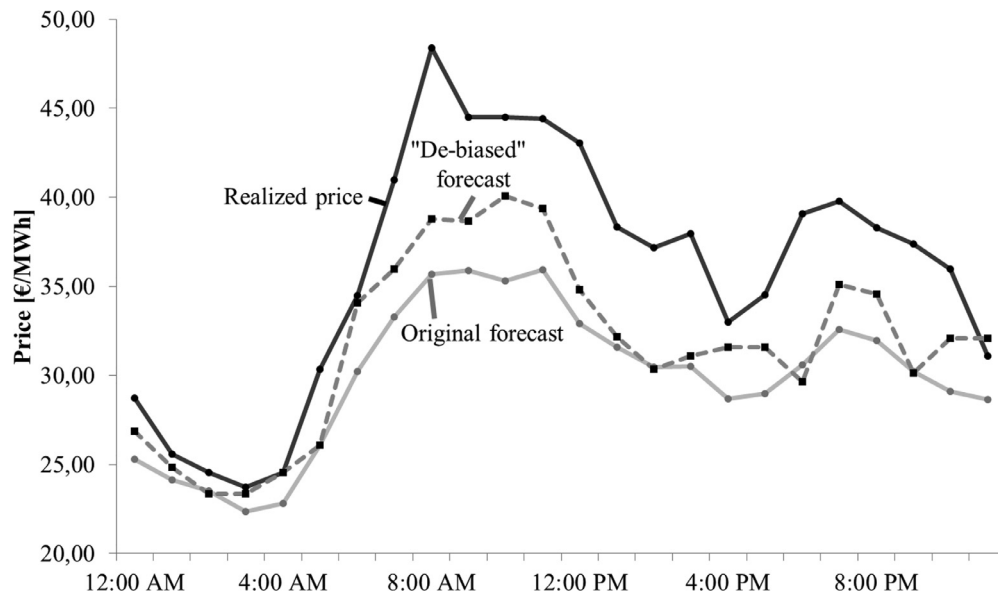

Fig. 6. Example of forecast de-biasing.

4. Proposed improvement actions

Several improvement actions were identified based on the findings of the previous section, free text commentaries and interviews of both planners and managers of the case company. Here, we describe two of those in more detail and present a summary of all recommended changes. Section 4.4 provides a follow-up of improvements.

4.1. De-biasing electricity price forecast

We first investigate the price forecast and its role in decision making. Theoretically, if inputs are correct and model captures all relevant aspects of a decision problem, manual adjustments to output should not systematically improve the result. But analysis in Section 3.4.3 shows that when planners make adjustments based on "Market" reasons, their actions seem to result in financial

benefits. As indicated in free-text answers, a typical reason for these deviations is the planner's distrust in the electricity price forecast $\hat{p}(t)$. Specifically, the planners reported that sometimes they did not trust the forecast's hourly profile and made adjustments, directly to the production plan, based on, e.g., yesterday's price profile or other intuition-based reasoning. This is somewhat expected, as in the spot market relevant to the case company, electricity price fluctuates, and even sophisticated forecasts can be inaccurate at times – for example, 20% MAPE for one day is not uncommon. The forecast used in optimization is provided by company's internal research team.

To evaluate forecast accuracy, we compared day-ahead forecasts against actual prices from a two-year period. The analysis reveals that for some weekdays and hours, there is systematic and statistically significant over- or under-forecasting. Due to the sensitivity of the forecasting data details have been omitted, but as an example during Monday morning hours (say, between 8AM and 10AM),
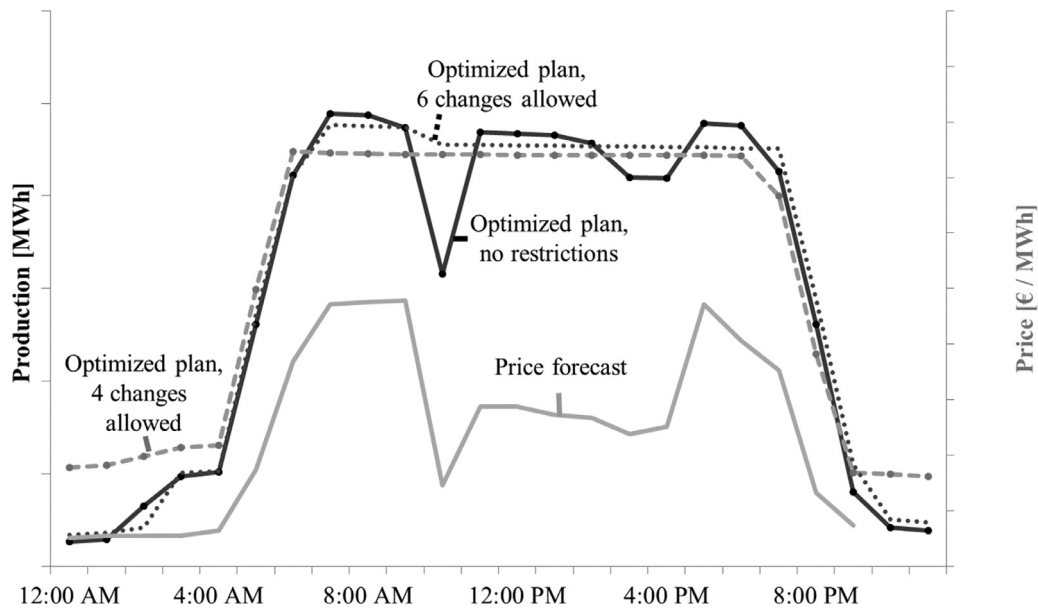
Fig. 7. Illustration of the impact of change restrictions to the plan. Note that the number of changes is restricted per plant, thus yielding a total plan with more than just 4 or 6 changes.

the forecast is systematically lower than the realized price: statistical one-sided *t*-test for difference between the actual price and forecast during 96 hours (32 Mondays × 3 hours) has $p < 0.01$. Then again, the difference between the actual price and forecast appears randomly distributed around zero at Saturday afternoon ($p > 0.1$).

The forecasting errors can be partially explained by the complexity of an auction-based electricity market; for example, in a review of hundred papers, Weron (2014) found that no single forecasting method has proven to systematically outperform others. A typical problem is underestimation of price peaks, which easily leads to systematic negative bias. But in the company case, also the forecasting process plays a role: for example, the analysts only work on business days and thus, Monday's forecast is done on Fridays, and so uncertainties pertaining to weather, for example, are higher. It was also noticed that in addition to providing a base forecast (which is used in the optimization tool) the analyst can also speculate about a different, low-probability price profile. So, the analysts' forecast is not a mean forecast of all scenarios, but rather his view of the most likely scenario.

The optimization model presented in Section 3.2, however, maximizes the average profit only if the price forecast is unbiased, i.e., $E[p(t)] = \hat{p}(t)$, where $p(t)$ is the random variable capturing electricity price of hour t. But, as outlined above, the forecast used by the planners is not completely unbiased. Forecast accuracy could be improved by working together with the internal analyst team, but we found out that biases can also be reduced algorithmically. In particular, we carried out a simple test by fitting a standard random forest model (e.g., Breiman, 2001) to actual prices that used the following inputs as independent variables:

- Analyst's day-ahead forecast $\hat{p}(t)$ (currently used in optimization)
- Historical prices $t - 24$ and $t - 168$
- Hour (e.g., "08–09") and weekday ("Monday")

Even though the analysis is not conclusive, we found that (i) the analyst's forecast is the key driver in explain actual prices but (ii) adjustments based on history and/or different times of week and day can indeed increase the accuracy, as exemplified in Fig. 6. So,

de-biasing the forecast has improvement potential as better forecast translates, at least to some extent, to better production plans.

4.2. Restricting hour-by-hour changes

Many of the observed deviations were explained by planners' willingness to restrict hour-by-hour changes in the plan, which became apparent in the survey results: this particular reason was mentioned in 9.6% of the days and also separately in the free-text comments. On one hand, avoiding any changes in running the plants can prolong equipment lifetime and make operating a plant less risky but, on the other hand, the company has estimated the impact to be rather small and hydropower plants are designed to tolerate such changes. But as noted in Section 3.4.3, these kinds of deviations did not typically lead to a severe drop in the financial performance of the actual plan. Thus, we set out to investigate systemically whether limiting the amount of changes by adding constraints to the optimization model in Section 3.2 would decrease the plan quality or not, and whether it would make the model computationally more complex.

Tracking the count of changes in the plan requires binary decision variables for each plant $\delta_i(t)$ that record if a change has happened. Then the number of changes at plant i can be limited to δ_i^{MAX} by introducing the additional constraints

$$q_i(t) - q_i(t-1) \leq M\delta_i(t) \quad \forall i, t$$

$$q_i(t-1) - q_i(t) \leq M\delta_i(t) \quad \forall i, t$$

$$\sum_t \delta_i(t) \leq \delta_i^{MAX} \quad \forall i,$$

where M is sufficient large constant coefficient. Fig. 7 illustrates how limiting the number of changes per plant impacts the optimized plan. In this case, the achieved price by the unrestricted plan is 50.3€/MWh. Correspondingly when six changes allowed the achieved price is 49.9€/MWh (−0.8%) and 49.1€/MWh (−2.5%) with four changes.

To find out if the example is representative or not, we analysed how restricting the number of changes at each plant during the first 24 hours of the plan would impact the planning performance

Table 4
Performance with and without restriction for changes in plan (during first 24 hours).

Changes per day	First 24 h of plan		Full planning period (7 days)	
	Forecasted revenue (%)	Actual revenue (%)	Forecasted revenue (%)	Actual revenue (%)
Unconstrained	100.00	100.00	100.00	100.00
6	99.95	99.94	99.99	99.99
4	100.04	100.05	99.99	99.98
3	100.25	100.27	99.96	99.94
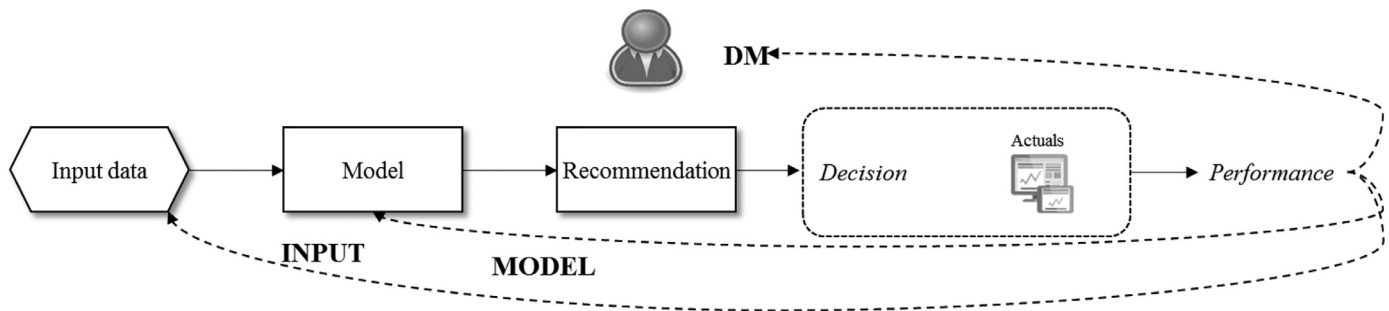
Fig. 8. Three categories of improvement actions.

over the entire study period. Unfortunately, it turned out to be impossible to fully capture how the hydro-balance would have developed throughout the 229 day period. This would have required extensive additional modeling efforts comparable to a separate development project. Thus, we instead replicated an average period regarding water availability and solved the 7-day optimization problem with forecasted prices for each day in the study period. The corresponding revenues were then calculated using both forecasted and realized prices. Results are in Table 4 below.

The results show that forecasted revenue for the whole planning period is highest when the number of changes is unrestricted (last two columns in Table 4). This is expected, as the objective of the planning model is to maximize expected profit and adding constraints can only decrease the objective. During the first 24 hours of the plan, which is the actual decision horizon, restricting changes can even appear beneficial (first two columns): with less possibilities to change production levels, the plants typically operate closer to efficiency curve maximum and produce more energy and, in some cases, also more revenue if price variation is low or forecasts inaccurate. In the whole planning horizon, however, stable plans in the first 24 hours imply decreased profits on following days mainly because in restricted cases more water is spent during the first day. In any case, the main finding of Table 4 is that changes are small, almost negligible, and they translate into an annual loss of thousands or tens of thousands euros at maximum.

This loss is relatively small because the overall annual revenue is tens of millions.

Thus, restricting the amount of changes does not significantly deteriorate the performance of the plan. Since the earlier results in Fig. 5 show that the manual adjustments under "Technical" category did not increase the Δp values in our case study, we conclude that including the presented constraints for managing the hour-by-hour changes in the model could be worthwhile, as it would implement the desired plan characteristics implied by planners' current behavior without sacrificing the financial performance. It should also be made possible for the planner to test different parameter values and observe the economic impact while planning. Even though the change in formulation converts the problem from LP to MILP, initial tests indicate that new model computation times remain reasonable, i.e., few minutes at most.

4.3. Other improvement actions

We classify improvement actions into three categories depending on how performance feedback impacts the different entities of the decision process (Fig. 8). All eight actions are detailed in Table 5.

The first action (1) is a fundamental one and proposes shifting the decision process from output-adjusting interaction to input-adjusting interaction according to in Fig. 1. We found that the

Table 5
List of identified improvement actions for the case company.

Object	Action	Motivation/potential benefits
Decision-maker	1. Modify inputs instead of outputs	Modifying inputs and re-running a model guarantees that all model dynamics are accounted for in recommendations (i.e. input-adjusting interaction)
	2. Enable automatic performance feedback	Planner does not see automatically the cost/value of his decision to change optimized plans; providing immediate automatic response enables objective self-assessment and promotes continuous on-the-job learning
Model	3. Improve hydro modeling	Improved accuracy in hydrological model would increase trust and decrease probability of deviations
	4. Limit adjustments per day	Planners need to restrict the number of changes within one day's plan
	5. Include hydrological risks in objective function	Allow control of planning parameters related to reservoir limits and penalty costs of braking these soft limits
	6. Improve price forecast	Price uncertainty is high and difficult to account for manually
Input Data	7. Improve and create new hydro data	Needed for better modeling of the underlying physical system

optimization model was not used iteratively by the planners in our case, i.e., the model is practically always run only once and adjustments are done directly to the final plan. However, in a complex system like the chain of hydropower plants, it is difficult for a human planner to foresee the effects of adjustments to the total plan: for example, decreasing output of one plant upstream can have a complex dynamic impact on the reservoir levels and production possibilities downstream. It can be argued that making adjustments related to any of the inputs (e.g., price forecast, plant minimum/maximum operating levels, reservoir level limitations) systematically and then re-running the model is likely to give better results. Another decision maker related improvement action (2) is promoting continuous learning through an automatic feedback reporting.

In this case, the planner feedback indicated that the model does not yield high quality plans in hydrological sense; e.g., reservoir levels do not behave as the model predicts. Hence, new ways to capture river dynamics with, e.g., dynamic lags dependent on flow level should be explored (3), but this would also require improving data and gathering new data by calibrating existing measurements and installing new meters for flow rate (7). As discussed in Section 4.2, the planners should have an option to restrict the number of changes in the plan (4). We recommend that the cost of this constraint is made visible to avoid extensive restrictions that limit profitability.

From the planners' feedback, it also became apparent that some hydrological situations are riskier than others, and that sometimes the optimization model produces too risky plans. This can be managed, for instance, by letting the planners control the penalty costs or soft limits for the reservoir minimum and maximum levels (5). Our fourth model-related improvement action proposal is improving the forecast (6), as explained in Section 4.1. Note that this can be quite impactful: the planners estimated high market uncertainty in over 30% of days, and they also outperformed the optimization model on those days when they had been doubtful about the price forecast.

The analysis helped identify improvement actions, the implementation of which is outside the scope of this study. For example, maintenance breaks (one key driver in the "External" category) can worsen the result of planning significantly. But improving maintenance scheduling by, e.g., jointly optimizing for hydropower planning and maintenance, is outside our scope. Similarly, some limitations to planning are due to unwritten rules which seek to maintain good relations with the inhabitants, especially those owning property on the shores of the affected lakes and rivers (lake and river coast). In practice, these rules might set limits to reservoir levels that tighter than those required by the environmental laws and other regulation. According to planners' feedback related to the "Technical" category, these often lead to small deviations that could perhaps be eliminated with the risk of reputation loss and other intangible harm.

4.4. Reflections on proposed changes

Two years after the study and its recommendations, we conducted interviews with the manager responsible of production planning and the expert responsible for the optimization tool development to follow up on realized improvements in the hydropower planning process. It is challenging to attribute changes done in the company to general business development actions or benefits of this study alone, not least because the operating environment and personnel have changed since the beginning of this study. However, some changes in both forecasting and planning model development can be tracked.

The forecasting of electricity prices is fundamental for decision making but at the same time complex from both technical and fundament perspective (Weron, 2014). As recommended, the case company has initiated continuous improvement in their internal forecasting process because higher forecast accuracy should lead to fewer adjustments due to planners' conflicting views. Since the company has multiple uses for the electricity price forecast, they have decided to centralize forecast improvement actions to an analyst team. Additionally, they have recently upgraded to a new tool for short-term forecasting, at least partially because of the requirements from hydropower planning as identified in this study (such as the need for improved accuracy and price scenarios).

The planning model has been revised to allow restrictions on the number of changes at plants but we do not possess numerical data to assess its performance. Then again, due to changes in its hydropower fleet, the case company is currently focusing on increasing automatization in both planning and operations which, in turn, might decrease the need to limit the amount of changes in the future if plants are operated by automation systems instead of human planners. It appears that the company is performing a cycle where the DSS was first developed by taking DMs' perspectives better into account, and then further developed towards fully automatic decision making. In this development, the role of planner is driven towards analyzing the results and external factors (e.g. opportunities in different electricity markets are not included in the model), whereas running the model will become more automated. The plan is to run the model constantly, so that up-to-date optimized plan will be available 24/7 for purposes other than the day-ahead trading. The feedback from the company managers confirms that model-driven decision processes require continuous improvement and interventions such as the one reported here can really help changes that are pertinent for the business environment at the time of a study.

5. Managerial implications

Implementation of model-based DSS essentially consists of design and deployment. There are several widely used frameworks that detail activities for this process, for example, INFORMS CAP Analytics Process or CRISP-DM. When decision support is in use its lifecycle activities should include monitoring and analyzing deviations between model recommendations and decisions. Observing such deviations ideally triggers the intervention process used in this paper and outlined in Fig. 9.

Based on our learnings, even the triggering of the intervention process might not take place because not all managers actively monitor how accurately model recommendations are followed. Other noteworthy observations from our study are the following. First and foremost, it is essential to cover relevant stakeholders (process owner, model developer, and DM) in the initial analysis phase (Step 1): in this way business logic and key assumptions can be cross-validated. At best, this phase also includes site visits and monitoring of the decision process in practice in addition to exploring existing data. In Steps 2–4, we found that collecting and combining objective and judgmental process feedback is extremely useful for recognizing improvement actions, in contrast to less structured approaches. But it is also critical to balance the trade-off between collecting judgmental data and intervening the DMs daily work: for example, filling out a daily survey can be effective but only if it has relatively little content and requires at maximum few minutes to fill in. We also found that it is very difficult to make adequate data specifications without observing any data, so therefore we recommend testing the study setting for a small period, running planned quantitative analyses, checking survey results and interviewing DMs to improve the study setup before initiating the actual data collection (Step 5). After the sanity check and possible changes in the study design, data collection with sufficient instructions can be triggered. It is
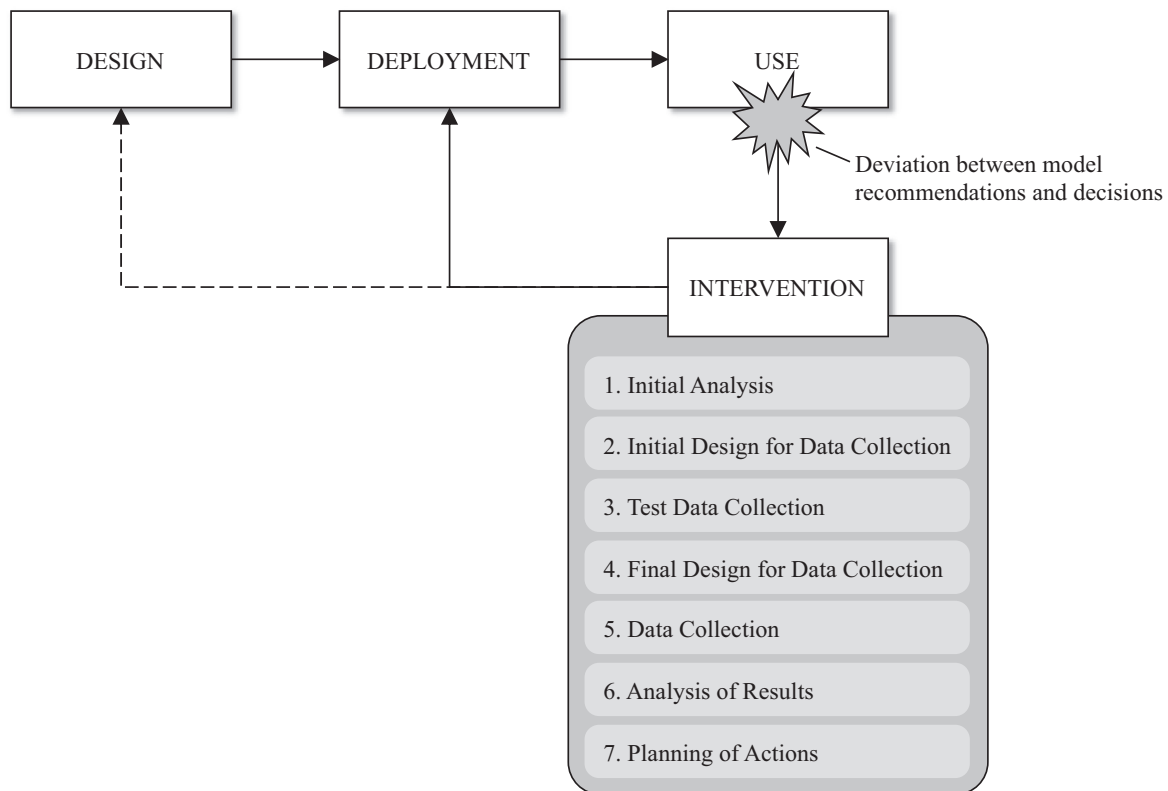
Fig. 9. The intervention process as a part of model-based DSS lifecycle.

important to monitor data accumulation to ensure that its content and quantity develop as planned. We partially failed in this respect, as we were not able to guarantee 100% fill rate for the survey. It is also essential not to interact with DMs during the study extensively, as this might contaminate the data by, e.g., making the DMs change their behavior based on discussions with researchers.

The analysis phase of Step 6 should result in objective findings about the process. In our study, the interviews in Step 1 indicated that distrust in the electricity price forecast is the key driver for the planners' deviations. In Step 6, we found that most deviations were done for reasons that are not related to the forecast at all as the judgmental feedback from daily surveys revealed that distrust in forecasting is just one-out-of-many problems in the process. We also found that a technical reason was the cause for deviations in 26% of the days but investigating the financial impact of the deviations on those days revealed that the impact is negative, but insignificant. Finally, Step 7 establishes a link between objective analysis and improvement actions that are realistic, considering organizational (e.g., change resistance), technical (e.g., computational complexity), and financial (e.g., budgetary) constraints. Examples of such actions can be found in Section 4 of this paper.

One finding deserves special attention. In Step 1 of our case study, we could not identify data to show why planners made changes to optimized plans, so those had to be surveyed. In an input-adjusting process, this would not need be the case. For example, if planners made changes to price forecast directly instead of implicitly adjusting the output, according to their beliefs about the forecast, the process could perhaps be sufficiently analyzed without asking the planners about their intentions. In the same vein, if planners would adjust the technical characteristics of plans using optimization constraints instead of changing the output, it would again be possible to isolate the financial effect explicitly without the survey method used in this paper. We have shown that rigorous analysis of output-adjusting process is possible but,

as a rule, implementing input-adjusting decision processes should make the whole intervention process smoother and richer in detail.

The findings of this study are generalizable to a wide array of decision processes. This is due to the characteristics of the planning process: the decisions (production quantities) are made under uncertainty (price) and the decision space in constrained (hydrological constraints) but still has flexibility. Similar properties are found in many decision processes such as inventory management, capacity planning, sales and operations planning, or project portfolio management.

6. Conclusions

The real-life use of decision models seems to receive less attention than theoretical model development even though the research on the value of modeling and implementations has long roots (Liberatore et al., 2000; Little, 1970). We have studied how DMs' (here, planners') decisions differ from the recommendations of a model-based DSS (in this case a hydropower optimization tool). The main result was that even though the planners make some adjustments that have a positive financial impact, they mainly worsen the financial performance of the production plan with their actions. To further understand the motives of planners, the results were cross-tabulated with judgmental process feedback. While external reasons (e.g., supervisor's order; plant maintenance) explain a lot of the negative impact, many of these differences are due to model inaccuracies, model simplifications, and challenges in the input data. We then presented justified improvement actions using the findings from the study. Some of these such as a new Mixed-Integer Linear Programming formulation of the model have already been implemented at the company.

This paper contributes to the literature by presenting a relevant study that, in contrast to the studies that use either objective data (e.g., Elmaghraby et al., 2015, Van Donselaar et al., 2010) or

judgmental feedback, utilizes both objective and judgmental feedback. This provides a great avenue for fact-based, effective intervention that aims at improving decision processes. Conceptually similar studies with different applications have been done, but in laboratory settings (e.g., Gavirneri & Isen, 2010).

The practical relevance of our study is significant because managers of almost any model-assisted decision process can adapt our intervention process and implement improvement actions proposed in the paper. Even though an increasing number of decision processes can be automated and improved (e.g., with machine learning), there are still numerous processes that will be dependent on a decision maker. Our study is particularly helpful in developing and improving such model-assisted processes internally, or with partners using the facilitated modeling approach. Our approach may not be well-suited for improving fully automated decision processes or strategic one-off decisions. We also acknowledge that collecting feedback data is not always a straightforward task, and in some cases comparing decision recommendations and actual decisions is impossible. It is also apparent that in order to utilize the framework, some investments into data collection and analysis are required, and in many cases these might not be economically justified.

There are multiple promising research avenues that stem from this research. First natural extension is to conduct more similar studies for different applications to understand better the motives of planners to deviate from model recommendations and identify improvement actions. To provide a starting point for such endeavors, we have listed potential improvement actions for model-driven decision processes to Appendix A, based on the case study and complemented by the earlier field experience of the authors. Second, for decisions done under uncertainty, comparing ex-post optimal decisions with realized uncertainties to ex-ante optimal recommendations and/or decisions could bring novel insights and result in a continuous process improvement framework. There is also demand for laboratory or real-life experiments of testing systematically how improvements (e.g., de-biasing forecast as proposed in Section 4.1) impact the DM behavior. Finally, there is an increasing need for research where machine learning is used for adaptive decision processes, in which DSS can learn from human contributions and correct human mistakes. This would lead to more automatic, seamless and, in general, better human-model interaction.

Appendix A. List of corrective actions

Table 6
List of corrective actions per three categories. Last column indicates whether external stakeholders are typically involved in performing the corrective action, in addition to an OR-expert.

Object	Corrective action	Potential benefits	External stakeholders
Decision maker	Change model/user interaction from output- to input-adjusting	Better utilization of model leads to systematically better plans without bias	No
	Align incentives	Faster and less biased decisions when model objective functions(s)/goal(s) are aligned with decision-maker's incentives	Yes (e.g., Management, HR)
	Establish continuous learning	Better and faster corrective actions in behavior, modeling or data, for example, in case of change in business environment	Yes (HR)
		Systematic scientific analysis of needed changes in human decision-making behavior	
	Develop performance indicators	Holistic and accurate indicators allow better comparison of recommendations and judgment	Yes (Management, HR)
		Promotes full use of model when results show its value	
	Improve analytics skillsets	Faster learning (e.g. through new hires and training)	Yes (HR)
		Better understanding of the model logic	
		Higher trust/commitment to model recommendations	
Model	Increase granularity	Fewer modifications to model recommendations based on decision-maker's own judgment	No
	Reduce complexity (if no impact on business benefits)	Better understanding of model logic	No
		Faster interpretation of model recommendations	
	Change objective function(s) / goal(s)	Fewer modifications to model recommendations (e.g. no need to consider secondary or implicit objectives/goals)	Yes (Management)
	Model uncertainties with robust/stochastic approaches	Higher trust in model recommendations	No
		Fewer modifications to model recommendations	
	Incorporate risk-attitude	Less need to manually modify inputs or recommendations	Yes (Management, Finance)
		Optimal risk mitigation	
	Add constraints to ensure implementable recommendations	No need for recurring basic modifications to input data/recommendation to obtain feasible recommendations	No
	Increase the performance of solution algorithms or computation infrastructure	More time for sensitivity/what-if analyses	Yes (IT, Data provider if external)
		Makes input-adjusting more tempting as iterative use of model gets faster	
Input data	Improve data quality (e.g. forecast accuracy)	More accurate model recommendations can even lead to process automatization	Yes (IT, Data provider if external)
		Possibility to automate	
	Monitor and update data (machine learning)	Less need for manual modification of input data/recommendations (automated and faster adaptation to environmental changes)	Yes (IT)
	Harness new data sources (objective/judgmental)	Better recommendations as a result of more realistic model parameters	No
	Capture data uncertainty (scenario/distribution-based forecasts)	Less need for human judgment; possibility to automate some risk assessments	Yes (Data provider if external)
		No need for offline risk assessment of recommendations	

References

Bolton, G. E., & Katok, E. (2008). Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management, 10*(3), 519–538.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5–32.

Durbach, I. N., & Montibeller, G. (2019). Behavioural Analytics: Exploring judgments and choices in large data sets. *Journal of the Operational Research Society, 70*(2), 255–268. doi:10.1080/01605682.2018.1434400.

Elmaghraby, W., Jank, W., Zhang, S., & Karaesmen, I. Z. (2015). Sales force behavior, pricing information, and pricing decisions. *Manufacturing & Service Operations Management, 17*(4), 495–510.

Flatabø, N., Doorman, G., Grande, O., Randen, H., & Wangensteen, I. (2003). Experience with the Nord Pool design and implementation. *IEEE Transactions on Power Systems, 18*(2), 541–547.

Fleten, S. E., & Kristoffersen, T. K. (2007). Stochastic programming for optimizing bidding strategies of a Nordic hydropower producer. *European Journal of Operational Research, 181*(2), 916–928.

Franco, L. A., & Hämäläinen, R. P. (2016). "Behavioural operations research: returning to the roots of the OR profession". *European Journal of Operational Research, 249*(3), 791–795.

Franco, L. A., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research, 205*(3), 489–500.

Fransoo, J. C., Waefler, T., & Wilson, J. R. (Eds.). (2010). *Behavioral operations in planning and scheduling*. Springer Science & Business Media.

Fransoo, J. C., & Wiers, V. C. S. (2006). Action variety of planners: cognitive load and requisite variety. *Journal of Operations Management, 24*(6), 813–821.

Fosso, O. B., Gjelsvik, A., Haugstad, A., Mo, B., & Wangensteen, I. (1999). "Generation scheduling in a deregulated system: the Norwegian case". *IEEE Transactions on Power Systems, 14*(1), 75–80.

Gavirneri, A., & Isen, S. (2010). Anatomy of a newsvendor decision: observations from a verbal protocol analysis. *Production and Operations Management, 19*(4), 453–462.

Hoch, S. J., & Schkade, D. A. (1996). A psychological approach to decision support systems. *Management Science, 42*(1), 51–64.

Hämäläinen, R. P., Luoma, J., & Saarinen, E. (2013). "On the importance of behavioural operations research: the case of understanding and communicating about dynamic systems". *European Journal of Operational Research, 228*(3), 623–634.

Kayande, U., De Bruyn, A., Lilien, G. L., Rangaswamy, A., & Van Bruggen, G. H. (2009). How incorporating feedback mechanisms in a DSS affects DSS evaluations. *Information Systems Research, 20*(4), 527–546.

Käki, A., Liesiö, J., Salo, A., & Talluri, S. (2015). Newsvendor decisions under supply uncertainty. *International Journal of Production Research, 53*(5), 1544–1560.

Liberatore, M. J., & Luo, W. (2010). The analytics movement: implications for operations research. *Interfaces, 40*(4), 313–324.

Liberatore, M. J., Hatchuel, A., Weil, B., & Stylianou, A. C. (2000). An organizational change perspective on the value of modeling. *European Journal of Operational Research, 125*(1), 184–194.

Little, J. D. C. (1970). Models and managers: the concept of a decision calculus. *Management Science, 16*(8), B–466.

Luoma, J. (2016). Model-based organizational decision making: a behavioral lens. *European Journal of Operational Research, 249*(3), 816–882.

Mild, P., Liesiö, J., & Salo, A. (2015). Selecting infrastructure maintenance projects with robust portfolio modeling. *Decision Support Systems, 77*(7), 21–30.

Montibeller, G., & von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis, 35*(7), 1230–1251.

Mortenson, M. J., Doherty, N. F., & Robinson, S. (2015). Operational research from Taylorism to terabytes: a research agenda. *European Journal of Operational Research, 241*(3), 583–595.

Nemhauser, G. L., & Wolsey, L. A. (1988). "Integer and combinatorial optimization". *Wiley-interscience series in discrete mathematics and optimization*. New York: Wiley.

Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research, 249*(3), 842–852.

Riedel, R., Fransoo, J., Wiers, V., Fischer, K., Cegarra, J., & Jentsch, D. (2010). Building decision support systems for acceptance. *Behavioral operations in planning and scheduling* (pp. 231–295). BerlinHeidelberg: Springer.

Riezebos, J., Hoc, J. M., Mebarki, N., Dimopoulos, C., Van Wezel, W., & Pinot, G. (2010). Design of scheduling algorithms. *Behavioral operations in planning and scheduling* (pp. 371–412). BerlinHeidelberg: Springer.

Schweitzer, M. E., & Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. *Management Science, 46*(3), 404–420.

Stoop, P. P. M., & Wiers, V. C. S. (1996). The complexity of scheduling in practice. *International Journal of Operations & Production Management, 16*(10), 37–53.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science, 185*(4157), 1124–1131.

Van Donselaar, K. H., Gaur, V., Van Woensel, T., Broekmeulen, R. A., & Fransoo, J. C. (2010). Ordering behavior in retail stores and implications for automated replenishment. *Management Science, 56*(5), 766–784.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research, 261*(2), 626–639.

Wallace, S. W., & Fleten, S. E. (2003). Stochastic programming models in energy. In A. Ruszczynski, & A. Shapiro (Eds.), *Handbooks in operations research and management science. Stochastic programming* (10, pp. 637–677). North-Holland.

Weron, R. (2014). "Electricity price forecasting: a review of the state-of-the-art with a look into t the future. *International Journal of Forecasting, 30*(4), 1030–1081.

Wäfler, T., Von der Weth, R., Karltun, J., Starker, U., Gärtner, K., Gasser, R., et al. (2010). Human control capabilities. *Behavioral operations in planning and scheduling* (pp. 199–230). BerlinHeidelberg: Springer.