Journal of Power Sources 413 (2019) 360–366

Contents lists available at ScienceDirect

Journal of Power Sources

journal homepage: www.elsevier.com/locate/jpowsour

Data mining in lithium-ion battery cell production

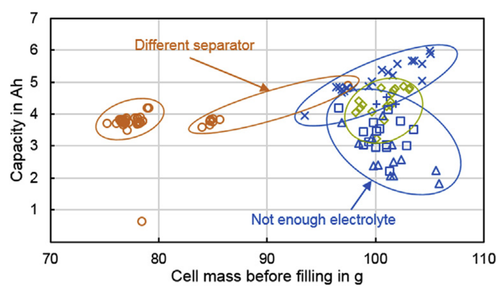
Joscha Schnell*, Corbinian Nentwich, Florian Endres, Anna Kollenda, Fabian Distel, Thomas Knoche, Gunther Reinhart

Technical University of Munich, Institute for Machine Tools and Industrial Management, Boltzmannstr. 15, D-85748, Garching, Germany

HIGHLIGHTS

- Data mining approaches were applied to a real battery production line.
- A systematic procedure for data acquisition, processing, and analysis is given.
- Electrode fabrication and electrolyte filling are identified as key quality drivers.
- The results can help to decrease battery production cost by reducing scrap rates.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Lithium-ion cell
Battery production
Data mining
Quality management
Process chain

ABSTRACT

Data mining methods are used to analyze and improve production processes in a lithium-ion cell manufacturing line. The CRISP-DM methodology is applied to the data captured during the manufacturing processes. Key goals include the identification of process dependencies and key quality drivers as well as the prediction of the product quality before the cumbersome and costly formation and aging procedure. Several Data mining methods, such as Generalized Linear Model (GLM), Artificial Neural Networks (ANN), Support Vector Regression (SVR), Decision Trees (DT), Random Forest (RF), and Gradient Boosted Trees (GBT) are compared and evaluated. Best results are yielded by an application of GLM, RF, and GBT for prediction of battery cell capacity before the expensive formation process. Key quality drivers identified are the electrode fabrication processes, as well as the electrolyte filling process during cell assembly. This is, to our knowledge, the first time data from a real battery production line has been systematically processed and analyzed along the whole process chain. The results of this paper can assist to manufacture better batteries and to reduce costs of lithium-ion cells by providing a systematic procedure for data acquisition and by lowering scrap rates during production.

1. Introduction

Regulatory standards for reduction of emissions and newcomers on the market have increased pressure on the automotive industry to provide affordable solutions for electric vehicles with sufficient driving distance and fast charging. The induced rise in demand for large-format lithium-ion cells will lead to expansion and new development of manufacturing capacities and increasing pressure to provide high quality cells at low production cost [1]. However, the complexity of the process chain for battery production [2] caused by the large number and variety of processes as well as unknown interdependencies between process parameters, intermediate product properties, and quality characteristics leads to high scrap rates [3] and immense effort for quality control [1]. Especially the cumbersome formation and ageing procedures before final quality check significantly contribute to the manufacturing costs [4]. Therefore, the identification of quality

* Corresponding author.
E-mail address: Joscha.Schnell@iwb.mw.tum.de (J. Schnell).

https://doi.org/10.1016/j.jpowsour.2018.12.062
Received 10 October 2018; Received in revised form 10 December 2018; Accepted 22 December 2018
0378-7753/

relevant parameters is crucial for cell manufacturers, plant engineering, and OEMs. While expert based methods for quality parameter identification have been successfully applied to the ramp-up of battery production facilities [2,5], fundamental evidence on the relevant interdependencies in battery manufacturing can only be gained by experimental validation [6–8] and data driven methods [9]. Due to the large number of processes and interactions along the process chain for battery cell production, a comprehensive experimental analysis of the production process will turn out to be challenging [9]. In contrast, data mining methods have already shown to be an efficient tool to improve manufacturing processes, for example in the semiconductor industry [10–12]. Similar methods were used to forecast end product quality for bio-chemical process chains, as well as for plastics production [13–15]. Huber et al. used data mining methods to categorize battery separator defects automatically by image recognition [16]. These approaches prove the applicability of data mining for quality assurance in complex process chains. However, to our knowledge, these methods have not been used so far to comprehensively analyze the influences of the whole lithium-ion cell production chain on final cell quality.

Therefore, the aim of this paper is to investigate the applicability of data mining methods in the fabrication of lithium-ion cells in order to identify key quality drivers and enable a prognosis of quality characteristics based on manufacturing data. The presented procedure aims to facilitate the capturing, processing, and analysis of data along the complete process chain by presenting "lessons learned" from a real pilot manufacturing line. The applicability of the method during ramp-up of production is investigated in order to assist battery cell manufacturers to implement procedures at an early stage of production, where only small amounts of data are available. Hence, efficient means of data processing and satisfactory data quality are ensured once production is running and larger amount of data are collected. This can help to understand and optimize production process and reduce scrap rates, promoting the production of better lithium-ion battery cells at lower cost. The structure of the paper follows the *Cross Industry Standard Process for Data Mining* (CRISP-DM) [17]: After an introduction to the pilot manufacturing facility used for data acquisition, the collection, analysis, and preparation of the data to be investigated are presented. Subsequently, different data mining approaches are compared regarding their applicability for quality parameter modeling. The created models are evaluated and the identified parameters are analyzed in a subsequent regression analysis.

2. Methods

Because of its holistic approach and its applicability to analyze data across different industries [18], the *Cross Industry Standard Process for Data Mining* (CRISP-DM) was chosen to analyze the data [17]. This method is divided into six highly iterative steps: Business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as shown in Fig. 1 (image a).

2.1. Business Understanding

Business Understanding describes the definition of overall goals to be achieved by data analysis in the respective business context. Derived from these goals, aims of the data analysis itself are determined and the initial situation of the DM context is evaluated [17]. The data base analyzed in this paper was generated in the pilot manufacturing facility for lithium-ion cell production of the Institute for Machine Tools and Industrial Management (*iwb*) at TUM [19]. The line can produce both hard case and pouch cells and consists of 19 consecutive process steps, as depicted in Fig. 1 (image b). Process steps 1–5 describe the electrode manufacturing from slurry mixing to calendering [6], which is carried through in a clean room. Afterwards the cells are assembled in a dry room in process steps 7–15 [20]. Process steps 16–19 describe the formation and final testing of the lithium-ion cells. During formation, the so called solid-electrolyte interphase (SEI) is built, which contributes significantly to the later performance of the battery cell. The packaging and sealing of the cells varies for the two different cell architectures, pouch cells and hard case cells, respectively. While the pouch bags are sealed from three sides (Step 13A) before electrolyte filling, a laser welding process is used for the hard case cells (Step 13B). The remaining side of the pouch cells is once sealed after electrolyte filling and a second time after degassing (Steps 15A and 17A), whereas the hard case cells are riveted in step 17B.

By the time of data acquisition, the data base contained 714 data sets (produced cells) from multiple production ramp-ups from different production batches with 1439 production and product parameters. Based on this information, the influence of process parameters on the quality of intermediate products, their influence on subsequent process steps and on the end product quality were investigated. The overall goal was to examine whether the quality of a cell can be assessed in an early state of cell production before the formation step.

2.2. Data understanding

In the *Data Understanding* step of the CRISP-DM-Method, the available data was reviewed, collected, relations were identified, and the quality of the data was evaluated. The production's data base [21] is organized in various tables, as depicted in Fig. 2. The data base consists of tables for different coils and slurries, reference tables for the anode and cathode coils, respectively, and tables for the assembled cell. Furthermore, information about the process parameters and quality checks for each produced cell and each process step are stored in tables. In the "Cell" table, one cell is listed per row and an ID for each process step is listed in the columns. These IDs can be utilized to identify the parameters used for the considered process and cell in the process tables.

The data for our analysis was collected during the ramp-up of different production batches. Thus, the analysis of data quality disclosed a wide gap between available and usable data. Since part of the electrodes were bought from commercial suppliers, the data sets for the electrode manufacturing were partly incomplete. Fig. 3 illustrates the amount of data sets for each process step during cell assembly with respect to the remaining amount of cells in the data base. A decrease by 81% of useable data sets with regard to the initial data set through all process steps was determined. In each process step manufacturing and monitoring data is collected. Manufacturing data (abbrev. "Process_" in Fig. 3) describes the process parameters used, while monitoring data (abbrev. "Monitoring_" in Fig. 3) consists of quality measurements during or after the process step. Especially the ultrasonic tab welding (process step 9) causes a decrease of 38% of useable data sets, presumably because of its high scrap rates. Therefore, this process step can be identified as a critical process step during ramp-up. Furthermore, only few or no data was stored for the process steps 10, 11, and 12 (fixation, drying, and positioning of the cell stack), as well as 13A, 15 and 17 (different sealing steps), mainly because these processes were carried through manually, process parameters were not varied or no quality measures were implemented for these processes. The main reasons for the overall loss of data along the process chain are the high scrap rate during ramp-up of production, the manual processes during cell assembly, the lack of sensors to satisfyingly collect data and, partly, the wrong or missing inputs in the data base by the users.

While for cell assembly (only steps 8–19) the remaining data set consists of 119 cells, the data set for the whole process chain, including data collected during electrode manufacturing (process steps 1–7), consisted of only 31 cells. The cells' quality attributes include in particular the cell capacity and internal resistance, measured in the testing procedure after formation [2]. Since the battery cells produced on the pilot line were used for different purposes after cell formation (i.e. safety tests, rate capability tests, cyclic aging etc.) with different test protocols, the data sets for cyclic aging were not comparable and could, therefore, not be included into the analysis.

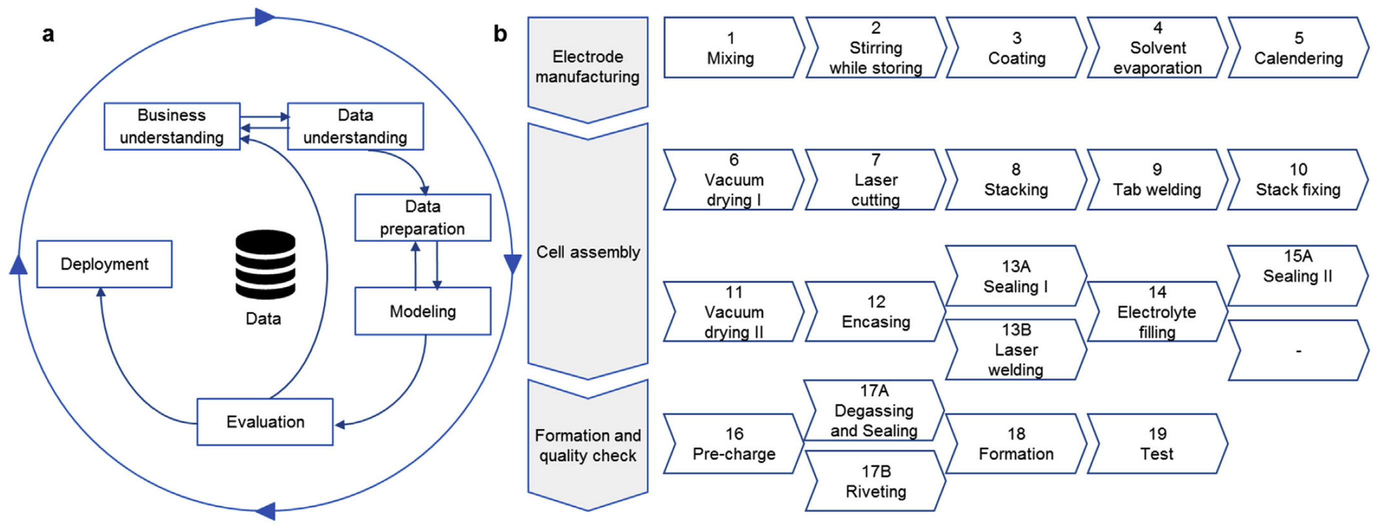


Fig. 1. CRISP-DM cycle following Chapman [17] (image a), process chain of the lithium-ion cell manufacturing facility used for data acquisition (image b). Different processes are carried through for pouch cells (A) and hard case cells (B) in steps 13, 15, and 17, respectively.

2.3. Data preparation

The next step in the CRISP-DM process is *Data Preparation*. Here, the usable data set is transformed into a processible data set by data selection, data clearance, data generation [22], data integration, and data formatting, as illustrated in Fig. 4. For this purpose, the initial set of parameters over all process steps was reduced from the data set to be analyzed by removing insufficiently recorded parameters of the electrode manufacturing as well as excluding parameters with no variance.

When selecting the relevant attributes, at first the structural characteristics such as the ID of the entry, the creator of the data record, the time of entry and the validity of the data record were removed. However, it must be ensured that no derived attributes are based on the parameters to be removed. Hence, implicitly given parameters such as the electrolyte mass or the wetting time of the cells had to be calculated before removal of the structural elements. For instance, the electrolyte mass could be calculated by the measured cell mass before and after the filling process step, while the wetting time is derived by the difference in time between the electrolyte filling step and the pre-charge. In the next step, all attributes that do not contain data or whose data has no variation were removed. This was done automatically using the "Remove Useless Attributes" element in Rapidminer [23].

Furthermore, non-numerical attributes (e.g. batch, separator type) were converted to numerical attributes. For this purpose, the different attribute values are represented as binary variables [22]. If, for example, the separator type is to be considered with the attributes "separator A" and "separator B", a new parameter set "separator A (binary)" and "separator B (binary)" was created, respectively. These new parameters were set to 1 if the respective cell had been assembled with the corresponding separator, or 0 in the other case. Then all text-based columns (e.g. comments, separator name, etc.) were removed. These columns were manually selected and controlled to avoid accidental exclusion of information. For example, some numerical values had been documented in text-based columns and, therefore, had to be manually transferred to newly generated numerical columns. When creating the individual analysis data records, other attributes that have a particularly unfavorable data situation were manually excluded. For example, attributes with more than 50% of all values missing were not considered by default.

The further processing of the data set included data clearance and data generation. Thereby, missing data was manually added where possible and reasonable and runaway values were evened by replacing them with mean values as proposed by Kantardzic [22]. One example for runaway values is the input of the measured resistance in mΩ instead of Ω. Since many of the runaway values were found for non-varied parameters, this approach does not cause any data corruption

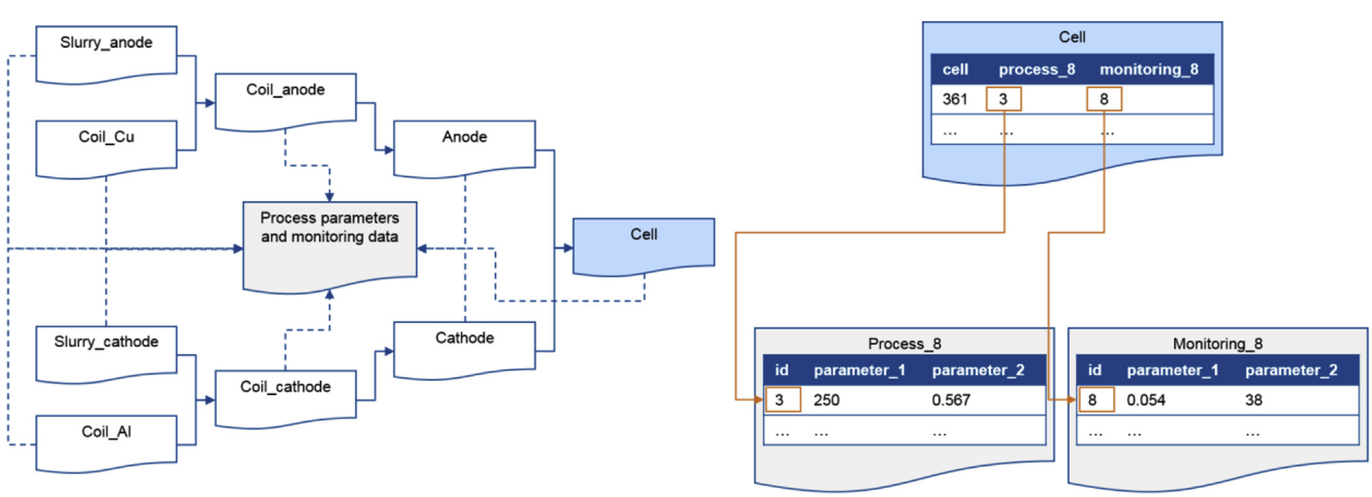


Fig. 2. Structure and referencing of the data base for electrode manufacturing and cell assembly, as suggested by Reinhart et al. [21].

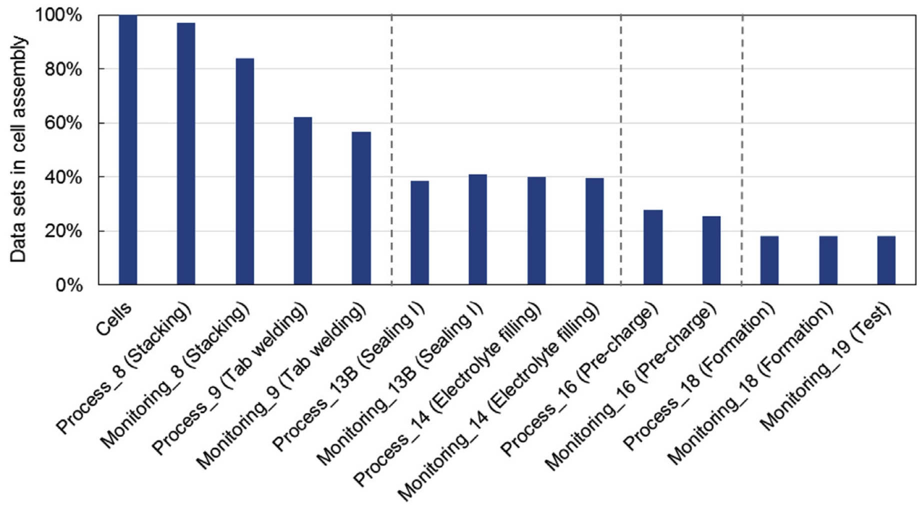


Fig. 3. Amount of data sets along the process chain in cell assembly. The horizontal gray bars indicate missing data sets of process steps 10–12, 13A, 15, and 17.

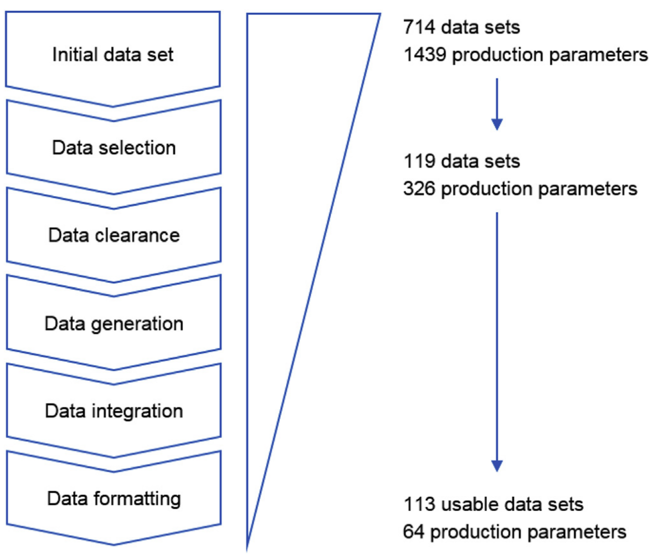


Fig. 4. Procedure for data preparation.

and, thus, is acceptable to keep a high number of data sets. Subsequently, the whole fragmented data set was merged and an excerpt of the data base was created to ease the processing of the data with the chosen data mining tool (RapidMiner).

In a final step, the data was normalized in order to ensure proper formatting and to prevent different parameter scales from distorting the final results of the analysis. Based on the assumption that the data are approximately Gaussian distributed, the data set was converted into a standard normal distribution with mean value 0 and standard deviation 1. The “Normalization” element integrated in RapidMiner was used for this purpose.

The largest data set obtained for the model consisted of 113 battery cells and 88 production parameters, mainly obtained during cell assembly. As the overall goal is the prediction of quality parameters before the time consuming formation procedure, data from pre-charge and formation was excluded for the prediction. Hence, the final data set comprises 64 parameters from electrode drying to electrolyte filling. The data contained six batches from multiple production ramp ups with different cell designs (number of electrodes in the cell stack) and cell chemistries (cathode materials $LiNi_{1/3}Mn_{1/3}Co_{1/3}O_2$, NMC and $LiFePO_4$, LFP).

2.4. Modeling

To analyze the prepared data set, different modeling techniques were compared, i.e. Generalized Linear Model (GLM), Support Vector Regression (SVR), Artificial Neuronal Networks (ANN), Decision Trees (DT), Random Forest (RF), and Gradient Boosted Trees (GBT). Basis for a quantitative comparison of the models in our analysis is the Root Mean Squared Error (RMSE), a standard quality measure for prediction of regression models which is used to compare the different modeling techniques [24]. It is calculated on basis of n data sets, the actual value y_i of the considered data point and the predicted value $f(x_i)$ based on the explanatory variables x_i. Besides the RMSE of the different modeling techniques, the RMSE of the mean value of the quality attributes was calculated to benchmark the prediction accuracy of the techniques:

$$RMSE = \sqrt{\sum_{i=1}^{n} (y_i - f(x_i))^2}$$

A ten-fold cross validation was used to increase the amount of training data [14]. Therefore, the whole data set was divided into ten equal-sized sub-sets. Each of these sets was used nine times to train the models in different combinations and the tenth data set was used to calculate the RMSE. This process was repeated ten times and the final performance of each model was evaluated by the mean of the ten RMSE values. A fully factorial experimental design was chosen to create the models by varying the model parameters, for instance the number and maximal depth of the trees in the RF and GBT models.

3. Results and discussion

3.1. Evaluation

Table 1 summarizes the results for the six different modeling techniques and the mean value with regard to the prediction of the cell

Table 1
Comparison of data mining tools for prediction of cell capacity before pre-charge and formation. The data set consists of 113 cells and 64 independent variables with an average capacity of 3.94 Ah.

Cell capacity	Mean	GLM	ANN	SVR	DT	RF	GBT
RMSE (in Ah)	0.89	0.42	0.76	0.48	0.51	0.44	0.42
Mean relative error (in %)		7.6	15	9.1	7.8	7.8	6.6
Improvement compared to mean (in %)		52	14	46	43	50	52

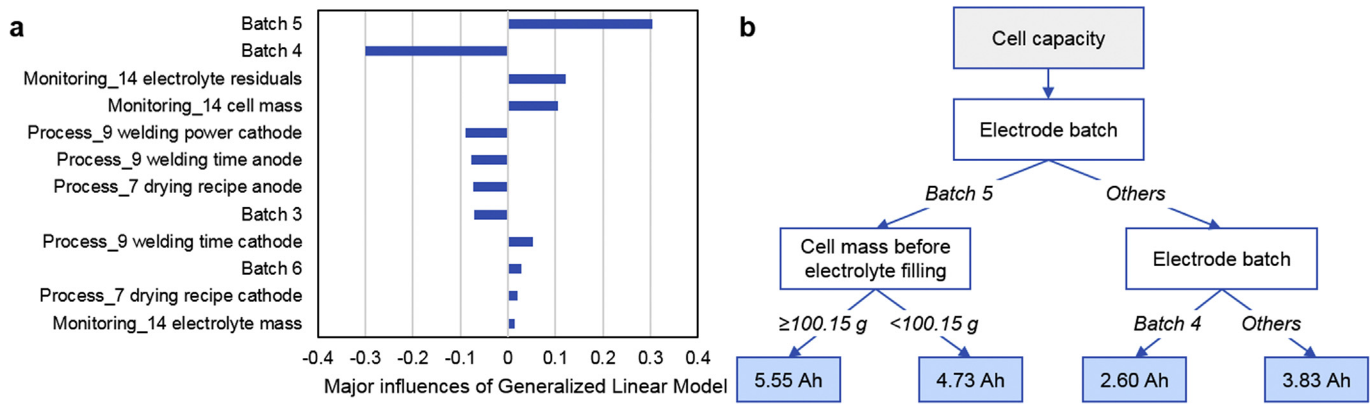


Fig. 5. Influences on cell capacity as derived from the GLM model (image a) and the DT model (image b).

capacity before the expensive formation step. Here, the smaller the RMSE, the better the prediction quality of the model. The column "Mean" describes the resulting model accuracy when simply using the average value to predict the respective quality parameter. For prediction of the cell capacity, the GLM and GBT models revealed the lowest RMSE, while the ANN model only showed slightly higher accuracy compared to the RMSE of the mean value. For the internal resistance, all models performed worse than the RMSE of the mean value and were therefore not further considered. An explanation could be that parameters with a significant influence on the internal resistance (e.g. the mixing process) were not included in the evaluated data set and, therefore, no correlation could be identified by the models.

Due to their easy interpretability, the GLM and DT models were further analyzed to identify the main quality drivers during cell production. Tracing back individual factors is hampered for the other models due to their black box nature (ANN and SVR [25]) or the large number of trees in the RF (60 trees) and GBT (20 trees), respectively. Fig. 5 (image a) illustrates the most important influences on cell capacity as derived from the GLM model. Unsurprisingly, the different electrode batches and the overall cell mass seem to have a strong influence on the cell capacity. Furthermore, the amount of electrolyte residuals after the cell filling process was identified by the GLM model. Further parameters derived from the GLM are mainly based on the ultrasonic welding process, e.g. the welding power or the welding time. This seems to be rather unintuitive, since these process parameters should not influence the cell capacity unless single layers were not properly bonded to the current collector. Fig. 5 (image b) shows the DT for the prediction model of the cell capacity, which only depends on two different electrode batches and the mass of the cell before electrolyte filling. A comparison of the two models reveals that apparently batch 5 has a positive impact on the overall cell capacity, whereas batch 4 results in rather poor capacities. The overall cell mass, as indicated by the GLM, is composed of the cell mass before electrolyte filling and the mass of electrolyte dosed during the filling step. Hence, the cell mass before filling will be taken into consideration in the further analysis (cf. section 3.2). Furthermore, the properties of the electrolyte filling process need to be carefully analyzed.

Of course, the models cannot provide information on process parameters which had not been varied or tracked during battery production. Therefore, it is highly likely that also parameters not listed here do have an influence on the final cell quality. Nonetheless, the presented approach shows that the modeling approach enables a determination of quality drivers in a real battery production line even with only a small data set (113 cells).

3.2. Deployment

To validate the models presented in Fig. 5, the correlations between capacity and attributes used for the GLM and DT were analyzed in detail and discussed with experts from the production line. Therefore the identified key parameters were plotted as a function of the cell capacity. As expected, no correlation could be identified for the ultrasonic welding parameters which had been predicted by the GLM. Fig. 6 provides an overview of the identified correlations in the GLM and DT model. The values belonging to batches with the same cell chemistry and target capacity (mass loading of the coatings, number of galvanic cells in the cell stack) are marked in the same color. Here, batch 1 was fabricated using an NMC cathode and a graphite (C) anode with a target cell capacity of 4 Ah. In contrast, batches 2 to 6 aimed at a target capacity of 5 Ah for the produced cells, i.e. more electrode and separator layers were stacked into these cells. Batch 2 contained an LFP cathode; the other batches also used NMC as the cathode material.

In Fig. 6 (image a), the cell capacity is plotted over the electrolyte residuals which, apart from the different production batches, had been identified as main influence parameter by the GLM. The electrolyte residuals are described by a discrete number between 1 and 5, where "1" corresponds to severe contamination of the cell housing with electrolyte residuals and "5" describes the highest optical inspection quality, i.e. no electrolyte outside the cell. The rating is the result of a qualitative (subjective) visual inspection based on worker experience. While for batches 1, 2, and 5, most cells were rated with "4" and "5" (i.e. only small amounts or no electrolyte residuals on the cell housing at all), batches 3 and 4 were mainly rated with values from "2" to "4". This means that during the electrolyte filling process a non-negligible amount of electrolyte could not be inserted into the battery cell, resulting in lower electrolyte distribution in the cell. Hence, in average these cells showed a significantly lower cell capacity than cells from the other batches.

This phenomenon is further consolidated in Fig. 6 (image b), which describes the behavior of the cell capacity in relation to the dosed electrolyte mass, as derived from the difference in cell mass before and after the electrolyte filling step. Here, batches 1 and 2 do not show any difference, while a clear trend can be observed for batches 3 to 6: Here, a positive correlation ($R^2 = 0.72$) is visible, i.e. the electrolyte dosed has a significant influence on the cell capacity. This behavior can be attributed to the distribution of electrolyte in the battery cell, where every pore in the electrodes and separator needs to be wetted to ensure ionic percolation [7,8,26]. If there is not enough electrolyte in the battery cell, part of the active material will remain "dry" and can, therefore, not contribute to the overall cell capacity. Hence, a saturation behavior is to be expected once sufficient amounts of electrolyte are inserted into the cell, which is apparently the case for batches 1 and 2. Note that batch 1 contains fewer electrode and separator layers, i.e. less electrolyte is required to achieve electrolyte saturation compared to the other batches.

In Fig. 6 (image c), the cell capacity is plotted as a function of the

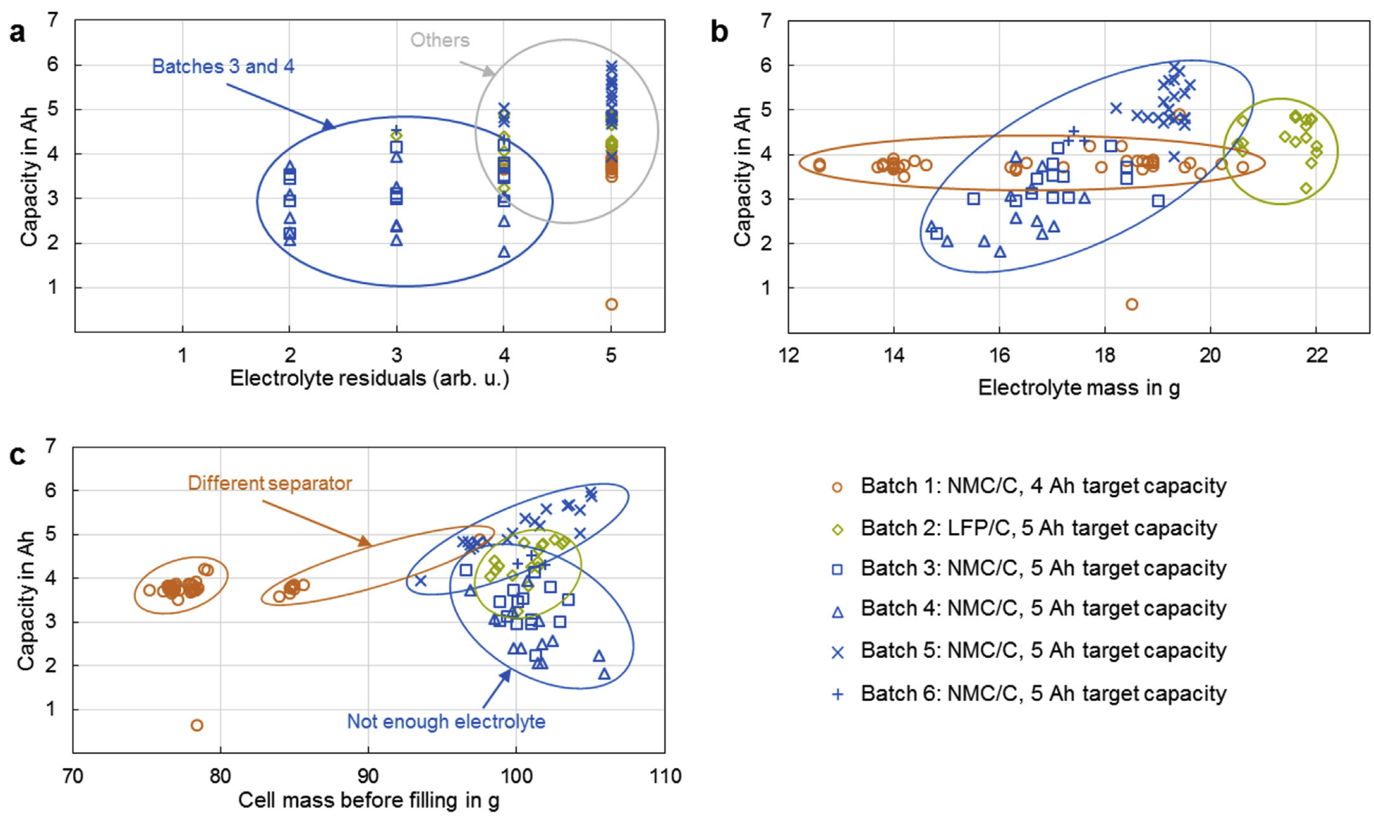


Fig. 6. Investigation of key quality drivers as deduced from the GLM and DT models. In a subsequent analysis of the corresponding data, the battery cell capacity was plotted as a function of the electrolyte residuals on the cell housing as derived by optical inspection ("5" corresponding to no residuals and "1" to severe contamination, image a), the electrolyte mass (image b), and the cell mass before electrolyte filling (image c), respectively. For an explanation of the indicated areas, please refer to the main text. Image c contains data already published in a previous paper [6].

cell mass before electrolyte filling, which was directly identified as a quality driver by the DT model and indirectly (via the overall cell mass) also by the GLM. Assuming that deviations in the mass of the inactive cell components (current collectors, separator, tabs, pouch bag) are negligible, the mass of the cell before electrolyte filling can be presumed as an indirect measure of the active material content in the cell. Therefore, a positive correlation with the cell capacity is to be expected. While for batch 5 (blue crosses), this positive correlation (coefficient of determination: $R^2 = 0.82$) is obvious, a more detailed investigation is required for the other batches. For batch 1, two different regions can be identified, one with a cell mass below 80 g and one with a cell mass larger than 84 g. Since, in average, both regions do not differ in cell capacity despite the difference in mass, the difference in mass needs to stem from one of the inactive components. Indeed, all cells in the region above 84 g were produced with a different separator, resulting in a higher cell mass but similar capacity. For batches 3 and 4, no clear correlation can be found, with a slight negative trend for heavier cells. This behavior can be traced back to the shortage of electrolyte in these cells: The larger the amount of active material is in the cell, the larger the surface and the SEI built-up during formation (Process 18), potentially resulting in a higher loss of active lithium and, thus, a lower overall capacity.

The results of the detailed parameter analysis confirm the strong influence of the electrode manufacturing processes (i.e. the different production batches) on the cell capacity [6], which was predicted by both the DT and the GLM model. Also the predicted influence of the mass before electrolyte filling could be confirmed using the regression analysis. Clearly, the model for the DT describes the relations between parameters and quality attribute in a strongly simplified manner, which might be an indicator for under-fitting. Nonetheless, this DT shows the same behavior as the results of the regression analysis presented by Günther et al. [6]. Although only predicted by the GLM model, the electrolyte filling process could be identified as a main quality driver, as earlier suggested in literature [7,8,26].

From these results it is obvious that expert knowledge is necessary especially for the final interpretation of the results. Clearly, data mining in production chains with a large number and variety of different processes is an interdisciplinary task which requires fundamental expertise both in data analysis, as well as the particular manufacturing processes. In this particular case, also knowledge of the electrochemical properties of lithium-ion cells is required. Obviously, only parameters actually traced and documented during production can be analyzed using the described approach. While the most significant parameters derived by the GLM and DT clearly provide reasonable results with regards to quality prediction, the models cannot be directly transferred to different production lines or, partly, even different production batches produced on the same production line: As both models strongly focus on the separation of different production batches (e.g., batch 5 and 4 as major influences in both the GLM and DT models), a new production batch will in most cases lead to different prediction results (over-fitting). Although investigations with a larger data set would be required for a comprehensive analysis and a higher prediction accuracy, the overall results impressively show that an identification of quality drivers in lithium-ion cell production is possible using data mining methods despite the small data base and the large number of independent variables. While the derived models can help to deduce the main quality drivers, the subsequent regression analysis can provide further valuable information on the correlations between process parameters and cell quality. This can help to improve production processes, reduce scrap-rates by setting appropriate process limits during manufacturing, and, thus, lower the overall cost of high-quality lithium-ion batteries.

4. Conclusion

In summary, data mining methods were analyzed concerning their applicability in lithium-ion battery cell production. The data collected during several production ramp-ups in a research production facility was processed on the basis of the CRISP-DM-Process. Therefore, data mining goals were defined and suitable data mining methods were selected. Using these models, the influence of process parameters on the end product quality was deduced. Best results were achieved for GLM, RF, and GBT for predicting the cell capacity. However, an easy interpretation of the resulting models is only possible for GLM and DT. The subsequent regression analysis showed that the electrode manufacturing, the mass of the cell before electrolyte filling and the amount of electrolyte in the cell have a significant influence on the cell capacity.

Several conclusions can be drawn from the presented results: First, it can be stated that data mining methods have been successfully applied in a real lithium-ion cell production line in order to identify process and monitoring parameters with a strong influence of the product quality. The biggest challenge for a comprehensive analysis along the entire process chain is to obtain a sufficiently large amount of complete data sets. For example, important attributes such as the slurry mixing parameters or the coating thickness had to be excluded from the analysis due to insufficient data quality. Additionally, parameters not varied during production were also not included in the analysis. Hence, vital prerequisites are the integration of measures to track process parameters and intermediate product properties, as well as the implementation of a user friendly data base covering all relevant process parameters [21]. While the pilot line used in this application scenario can only provide a limited amount of data due to the small output and the frequent ramp-ups with different material systems, established lithium-ion cell manufacturers could generate a large amount of data during production, although highly unlikely to share this kind of information. The methods presented in this paper could further be used to forecast intermediate product attributes and, thus, to identify parameters worth monitoring for quality assurance. Alternatively, the influence of intermediate products' characteristics on subsequent processes could be investigated as well. As earlier suggested in literature [9], a quality gate system could assist to aggregate all relevant data. Further refinement of the models and a bigger and more detailed data set are prerequisites to validate the presented results. The presented method could also be applied to monitor the energy flux along the process chain, helping to produce more sustainable batteries by lowering the energy consumption during manufacturing. Overall, the findings of this paper can assist in reducing scrap rates and, thus, producing battery cells with a higher quality at lower cost.

Declaration of interest

None.

Acknowledgements

We would like to thank the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) for the funding of our research. The results presented in this paper have been achieved within the scope of the projects "ExZellTUM" (FKZ: 03X4633A) and "ExZellTUM II" (03XP0081). We would also like to thank the whole battery production team, namely Nicolas Billot, Till Günther, Jakob Kurfer, Edgars Locmelis, Johannes Schmalz, Franz Spingler, Georgios Theodossiadis, and Markus Westermeier, for running the production line and, thus, generating the data used in the evaluation. Our special gratitude goes to Till Günther for proof reading of the manuscript.

References

[1] A. Kwade, W. Haselrieder, R. Leithoff, A. Modlinger, F. Dietrich, K. Droeder, Nat. Energy 3 (2018) 290–300.
[2] M. Westermeier, G. Reinhart, T. Zeilinger, 2013 3rd International Electric Drives Production Conference (EDPC), IEEE, 2013, pp. 1–10.
[3] A. Dinger, R. Martin, X. Mosquet, M. Rabl, D. Rizoulis, M. Russo, G. Sticher, The Boston Consulting Group, (2010).
[4] D.L. Wood, J. Li, C. Daniel, J. Power Sources 275 (2015) 234–242.
[5] M. Westermeier, G. Reinhart, M. Steber, Procedia CIRP 20 (2014) 13–19.
[6] T. Günther, N. Billot, J. Schuster, J. Schnell, F.B. Spingler, H.A. Gasteiger, AMR 1140 (2016) 304–311.
[7] T. Knoche, F. Surek, G. Reinhart, Procedia CIRP 41 (2016) 405–410.
[8] T. Knoche, V. Zinth, M. Schulz, J. Schnell, R. Gilles, G. Reinhart, J. Power Sources 331 (2016) 267–276.
[9] J. Schnell, G. Reinhart, Procedia CIRP 57 (2016) 568–573.
[10] H. Tsuda, H. Shirai, O. Takagi, R. Take (Eds.), Proceedings of ISSM2000. Ninth International Symposium on Semiconductor Manufacturing (IEEE Cat. No.00CH37130), Ultra Clean Soc, 2000, pp. 249–252.
[11] S.-C. Hsu, C.-F. Chien, Int. J. Prod. Econ. 107 (2007) 88–103.
[12] R. Evans, M. Boreland, IEEE 42nd Photovoltaic Specialist Conference (PVSC), IEEE, 2015, pp. 1–5 2015.
[13] Z. Cheng, X. Liu, J. Appl. Polym. Sci. 132 (2015).
[14] S. Charaniya, H. Le, H. Rangwala, K. Mills, K. Johnson, G. Karypis, W.-S. Hu, J. Biotechnol. 147 (2010) 186–197.
[15] G.T. Jemwa, C. Aldrich, AIChE J. 51 (2005) 526–543.
[16] J. Huber, C. Tammer, S. Krotil, S. Waidmann, X. Hao, C. Seidel, G. Reinhart, Procedia CIRP 57 (2016) 585–590.
[17] P. Chapman, CRISP-DM 1.0: Step-by-step Data Mining Guide, SPSS, 2000.
[18] G. Mariscal, Ó. Marbán, C. Fernández, Knowl. Eng. Rev. 25 (2010) 137–166.
[19] G. Reinhart, T. Zeilinger, J. Kurfer, M. Westermeier, C. Thiemann, M. Glonegger, M. Wunderer, C. Tammer, M. Schweier, M. Heinz, G. Schuh, R. Neugebauer, E. Uhlmann (Eds.), Future Trends in Production Engineering: Proceedings of the First Conference of the German Academic Society for Production Engineering (WGP), Berlin, Germany, 8th-9th June 2011, Springer, Berlin, New York, 2013, pp. 3–12.
[20] J. Kurfer, M. Westermeier, C. Tammer, G. Reinhart, CIRP Ann. 61 (2012) 1–4.
[21] G. Reinhart, J. Kurfer, M. Westermeier, T. Zeilinger, AMR 907 (2014) 365–378.
[22] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, second ed., Wiley IEEE Press, Hoboken, New Jersey, 2011.
[23] RapidMiner, https://rapidminer.com/, accessed 9 October 2018.
[24] T. Cao, E.-P. Lim, Z.-H. Zhou, T.-B. Ho, D. Cheung, H. Motoda, Advances in Knowledge Discovery and Data Mining, Springer International Publishing, Cham, 2015.
[25] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Mach. Learn. 46 (2002) 389–422.
[26] W.J. Weydanz, H. Reisenweber, A. Gottschalk, M. Schulz, T. Knoche, G. Reinhart, M. Masuch, J. Franke, R. Gilles, J. Power Sources 380 (2018) 126–134.