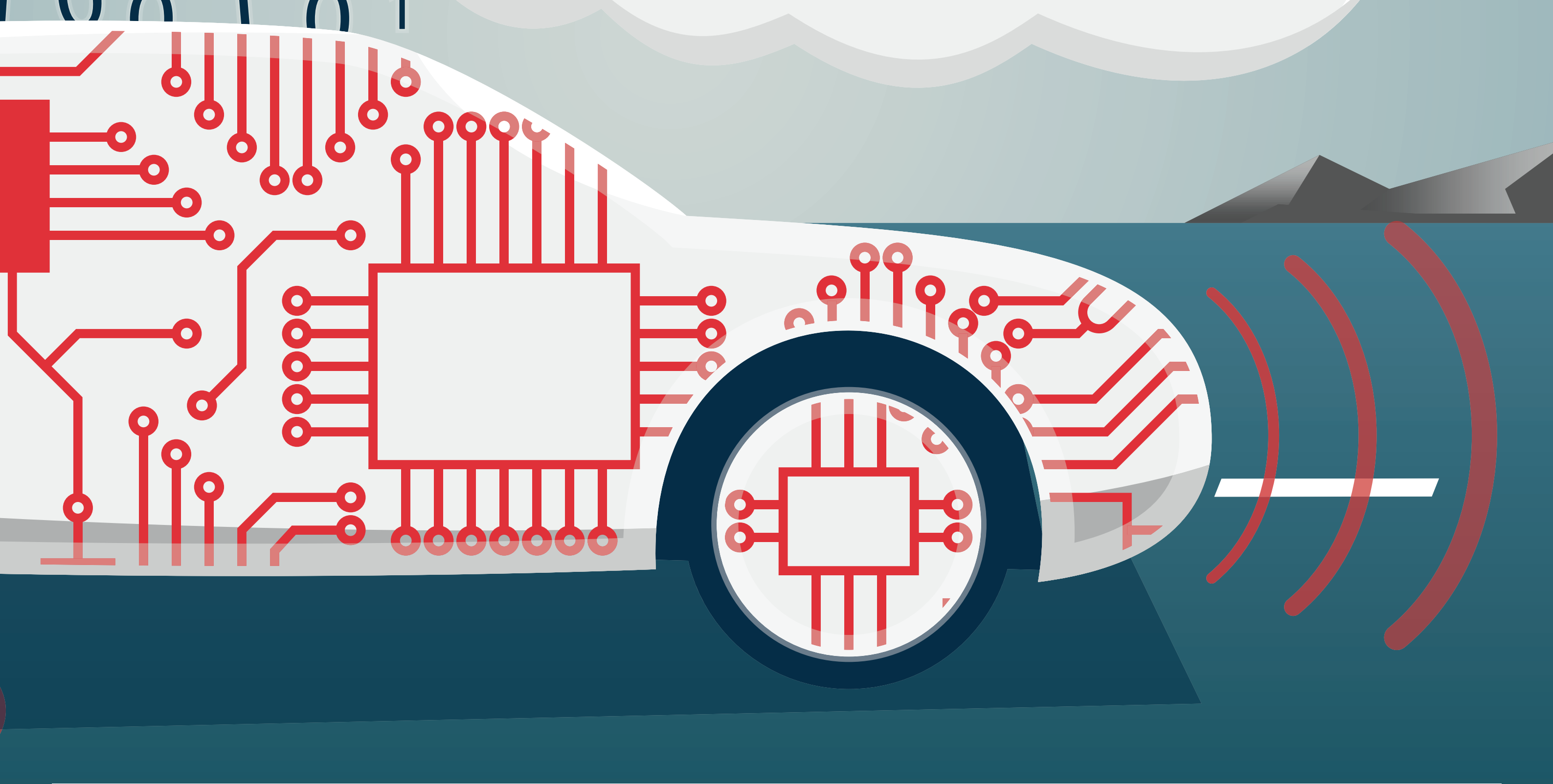
Springer Vieweg

ATZ extra

August 2018

120 YEARS

Big Data
Between Vehicle and Backend



Managing Big Data

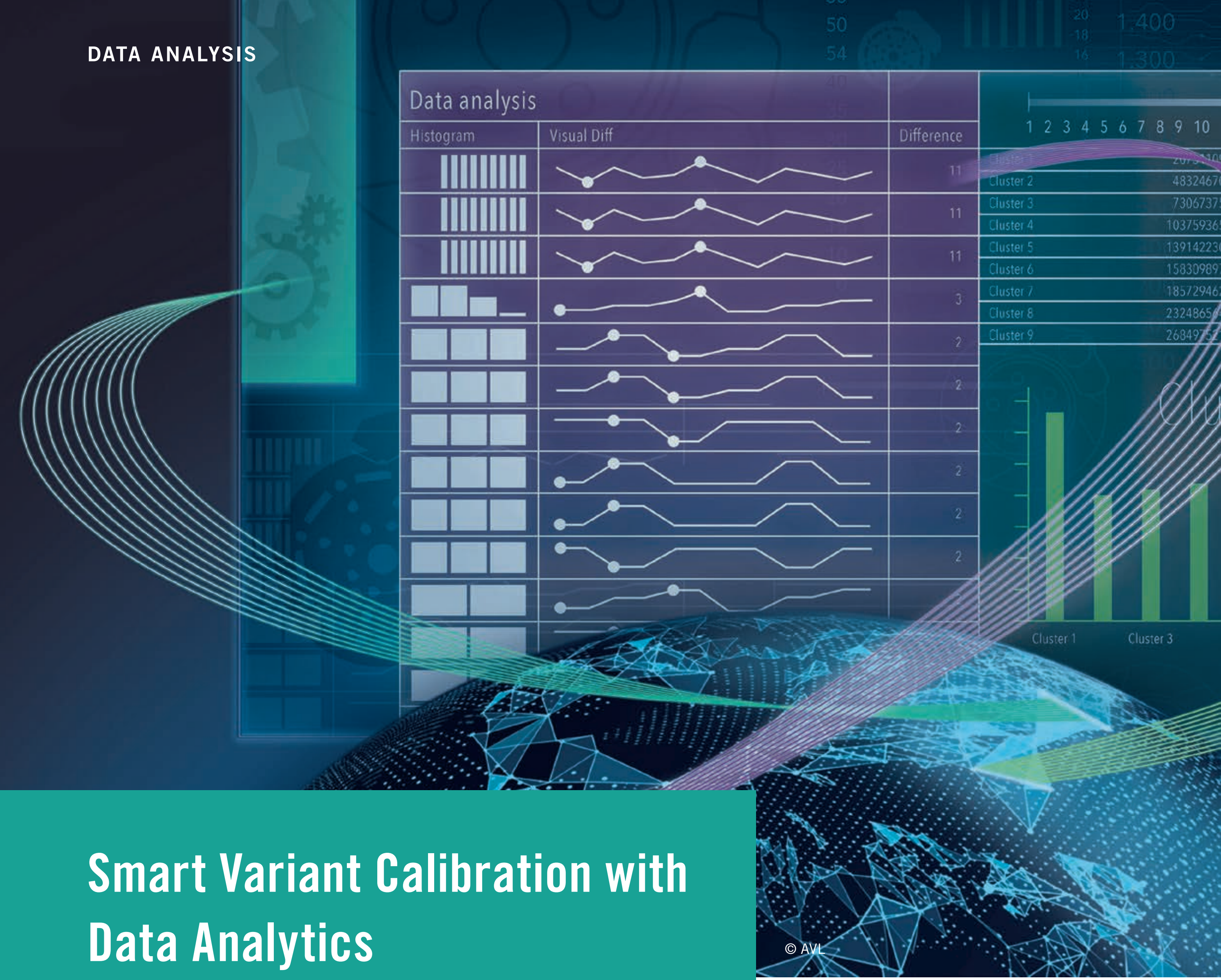
Dear Reader,

the rising number of mechatronic systems and the increasing connectivity of powertrain, chassis, and driver assistance functions are leading to a relentless increase in the complexity of calibration work on modern vehicles. As a result, we are seeing a dramatic rise in the number of labels in the ECUs. One of the challenges facing calibration engineers is how to manage this "big data" in the context of a growing number of vehicle variants and with calibration teams that are often located at different sites throughout the world. Only with a high level of efficiency and agility in the data management process will it be possible to implement these extensive calibration tasks in an increasingly shorter period of time and to meet the high requirements regarding functional safety, reliability, and robustness.

With its calibration data lifecycle management system AVL CRETA™, AVL has been the market leader in the field of calibration data management for many years now. The current fifth generation supports an agile calibration process by providing a wide range of features. For example, it ensures easy handling of the calibration during the vehicle development process, comprehensible integration of the ECU parameters into all projects, a clear representation of the calibration maturity, and conflict-free fusion of different data sets. Data management is optimized by powerful search functions, easy distribution of projects to global calibration teams, clear navigation through the different calibration variants, and advanced data mining algorithms. In total, this results in a time and cost reduction of more than 50 %. The software developers placed special emphasis on the simple re-use of calibrations for future applications. In the following article, experts from AVL present the latest component of AVL CRETA 5™, a tool for smart and efficient variant calibration based on data analytics methods. I hope you enjoy reading this interesting report.

Richard Backhaus

Richard Backhaus
ATZ correspondent



© AVL

Smart Variant Calibration with Data Analytics

The calibration of modern engine control units generates huge quantities of data. This data is not only a challenge for data management – it also includes a lot of knowledge that has remained unused up to now. AVL shows how this knowledge can be utilized for the efficient and robust calibration of variants using data analytics methods.

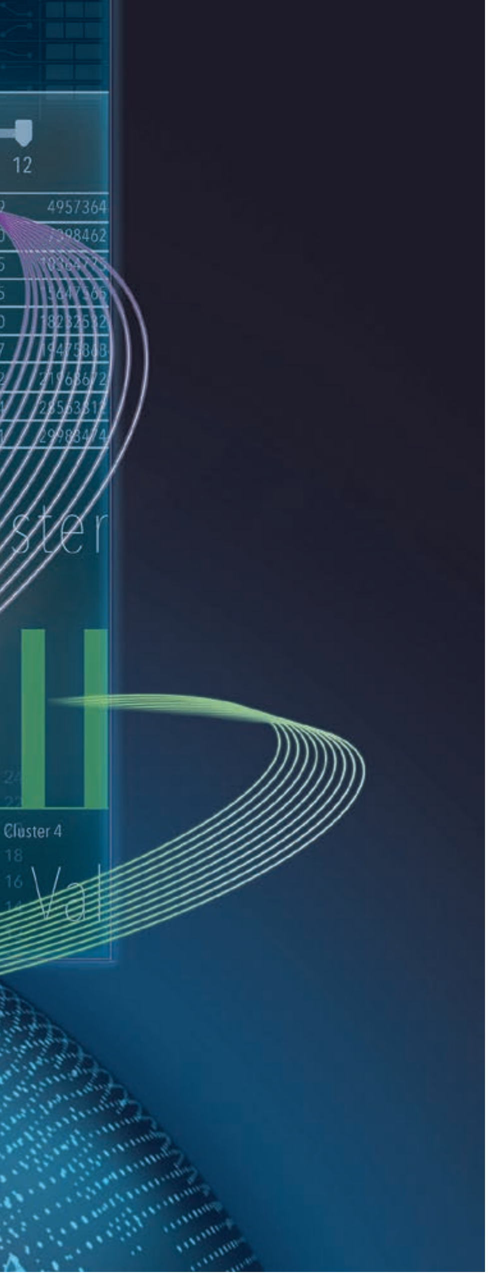
AUTHORS

Dipl.-Ing. Thomas Dobes
is Skill Area Manager for Gasoline Development & Calibration in the field of Powertrain Engineering at AVL List GmbH in Graz (Austria).

Dipl.-Ing. (FH) Thomas Kaserer
is Technical Expert for Data Management and Calibration Pilot in the field of Powertrain Engineering at AVL List GmbH in Graz (Austria).

Dipl.-Ing. (FH) Nikolas Schuch
is Head of the Department for Calibration Technologies in the field of IODP (Integrated Open Development Platform) at AVL List GmbH in Graz (Austria).

Gerhard Storfer
is Product Manager for AVL CRETA in the field of IODP (Integrated Open Development Platform) at AVL List GmbH in Graz (Austria).



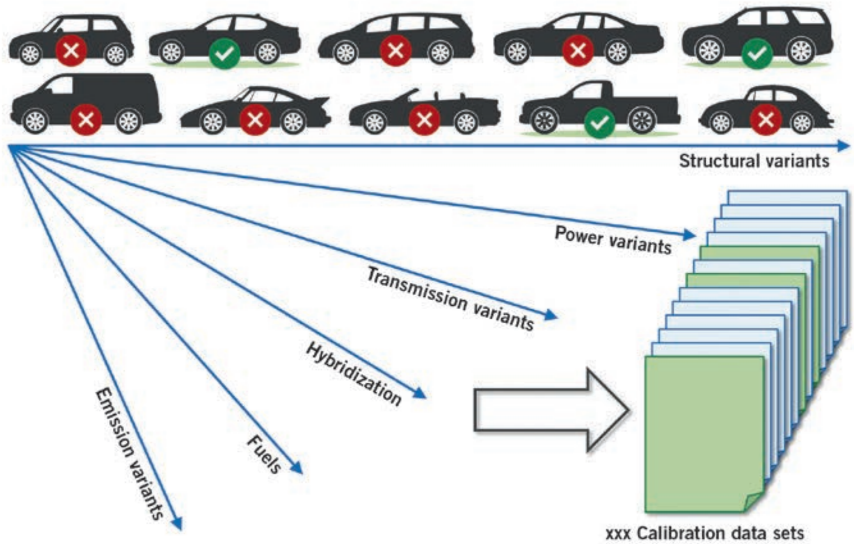


FIGURE 1 Calibrations are validated on lead variants (shown in green here) (© AVL)

Attributes/variants	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8	Variant n
Power	140	140	140	185	185	185	220	220	220
Country	EU	USA	CHN	EU	USA	CHN	EU	USA	CHN
Emission class	EU6d	LEVIII SULEV	CN5b	EU6d	LEVIII SULEV	CN5b	EU6d	LEVIII SULEV	CN5b
Fuel	ROZ95 ROZ98	CARB Ph.3	ROZ91	ROZ95 ROZ98	CARB Ph.3	ROZ91	ROZ95 ROZ98	CARB Ph.3	ROZ91
Start/stop	yes	no	yes	yes	no	yes	yes	no	yes
Displacement	1789	1789	1789	2155	2155	2155	2155	2155	2155
Chassis type	Limo	Limo	Station	Station	Station	Station	Limo	Station	Station
Supercharging	N. A.	N. A.	N. A.	Turbo	Turbo	Turbo	Bi-turbo	Bi-turbo	Bi-turbo
Engine mounting position	Longit.	Longit.	Longit.	Longit.	Longit.	Longit.	Longit.	Longit.	Longit.

FIGURE 2 Attributes of different vehicle variants (© AVL)

DATA QUANTITIES

50,000 labels per engine control unit, one million individual values, hundreds of vehicle variants, new data sets every week for 18 months – the data quantity generated in modern calibration projects has now become so large that it is de facto no longer possible to determine each label of the individual variants by experimental means. It is therefore common practice to define lead variants which are as representative as possible and to take over about 80 to 90 % of the labels from these lead variants. This is possible thanks to the high degree of synergy between the derivates. However, the crucial question is that as to which values represent these 80 to 90 %. Or to rephrase the question: Is it really necessary to invest time and money in order to safely reach the target values of the specification while ruling out errors?

For reasons of time and cost, only a few representative vehicles are now actually assembled to form real prototypes. These are used to represent all variants during final validation and approval, **FIGURE 1**.

Data errors in variants which are not available in reality would be fatal because they would no longer be discovered during the approval process.

Experts usually ensure that the focus of calibration work is laid on the critical parameters. Thanks to their long years of experience and their technical know-how, they are aware of the connections between the functions, parameters, and vehicle derivates, and this allows them to make the decisions necessary for the work of the calibration teams. However, what happens if even the expert reaches his limits? How can we simplify and safeguard his responsible work and optimize it if necessary? How can errors be avoided? And how can new connections be identified?

DATA ANALYTICS APPROACH

For AVL List GmbH, the solution is to be found in the intelligent analysis of current and historical calibration data. Patterns can be identified in this huge data quantity, and these in turn allow rules for variant calibration [1, 2] to be derived.

For this purpose, the approach starts from the technical attributes of the vehi-

Identical calibrations

Attributes/variants	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8	Variant n
Attribute A "Power"	140	140	140	185	185	185	220	220	220
Attribute B "Emission"	EU	USA	CHN	EU	USA	CHN	EU	USA	CHN
Variable 1									
Variable 2	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
Variable 3									
Variable 4	2.12		2.12	2.12		2.12	2.12		2.12
Variable 5	1.001			1.001			1.001		
Variable 6	12	12	12	12	12	12	12	12	12
Variable 7	0.101	0.101	0.101	0.101	0.101	0.101			
Variable n									

FIGURE 3 Template for the identical calibration of different variants (© AVL)

Identical calibrations

Attributes/variants	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8	Variante n
Attribute A "Power"	140	140	140	185	185	185	220	220	220
Attribute B "Emission"	EU	USA	CHN	EU	USA	CHN	EU	USA	CHN
Variable 1									
Variable 2	1.001	1.001	1.001	1.001	1.001	1.011	1.001	1.001	1.001
Variable 3									
Variable 4	2.12		2.12	2.12		2.12	2.12		2.12
Variable 5	1.001			1.001			1.001		
Variable 6	12	12	12	12	12	12	12	12	12
Variable 7	0.101	0.101	0.101	0.101	0.101	0.101			
Variable n									

FIGURE 4 Identification of outliers (© AVL)

cle variants, **FIGURE 2**, and links the attributes with certain patterns in the calibrations. This allows the identification of correlations between the data sets to help those responsible make reliable decisions for robust variant calibrations.

For the mass evaluation of the calibration data sets, AVL developed a data analytics tool that is made available within the calibration data lifecycle management system AVL CRETA [3]. This is used by OEMs and suppliers throughout the entire powertrain or vehicle development process for the central saving, management and merging of calibration data [4].

To allow the effective evaluation of this valuable database, the attributes must first be determined by the data manager or calibration expert. Experience shows that there are about 100 attributes with which the individual derivates can be characterized. Most of these attributes are already known from the calibration assignment formulated at the beginning of the project.

TEMPLATES AS A TOOL

The software development described here compares the data sets and derives suggestions for similarities and differences in calibration from the attributes and data patterns calculated. These suggestions can be imagined as templates allowing calibrations to be applied to other data sets. **FIGURE 3** illustrates the principle of a template of this kind. This example of a template highlights all identical calibrations of the variants in a project, thus allowing the creation of a rule stating that variable 4 must be identical for all EU and China variants. Another possible derivation would be that the calibration of variable 2 for the power variants with 140 hp, 185 hp and 220 hp must be the same in all national variants shown.

The templates identified must be validated before use. This prevents the creation of rules from random data patterns which are meaningless from a technical point of view.

If the expert approves a template, it can be used to derive the data set for derivates from the lead variants. The template can go on identifying potential calibration errors throughout the entire duration of the project. **FIGURE 4** shows an example of a deviation from the iden-

Attributes/variants	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8	Variant n
Attribute A "Leistung"	140	140	140	185	185	185	220	220	220
Attribute B "Emission"	EU	USA	CHN	EU	USA	CHN	EU	USA	CHN
					Attribute A				
Variable 1									
Variable 2									
Variable 3	101	101	101	102	102	102	103	103	103
Variable 4									
Variable 5									
Variable 6									
Variable 7	0.101	0.101	0.101	0.101	0.101	0.101	0.111	0.111	0.111
Variable n									

FIGURE 5 Template for compulsory calibration differences (© AVL)

tical calibration rule that has been highlighted in color. If it turns out that this is not an error but a compulsory difference, an additional attribute can be introduced. In other words, the models can be constantly improved in the course of a project.

Besides rules for identical calibration, it is also possible to derive rules for compulsory differences. **FIGURE 5** highlights the compulsory differences between three power variants. These could for example be the calibrations under full load of different power variants, which must therefore all be different too. **FIGURE 5** also shows an example in which compulsory differences are only required for a certain power class. This is for example the case if the higher power for these variants is achieved through a different hardware, for example another turbocharger. As a consequence, the relevant variable must be calibrated differently here.

The examples illustrate that the templates not only map mathematical correlations but also technical causalities. In this way, the valuable knowledge of the experts is conserved and made usable for other calibration engineers.

PROCEDURE MODEL

The data analytics process is based on CRISP-DM (Cross-Industry Standard Process for Data Mining). This widespread process model for data mining projects consists of six steps; feedback to previous steps is expressly intended, **FIGURE 6**.

STEPS 1 AND 2: UNDERSTANDING THE ENVIRONMENT AND THE DATA

A data warehouse obtains the data from the calibration data management system described above. It only obtains data which is especially meaningful and necessary for the data analytics algorithms. In addition to this information, it is also possible to include data from other sources, for example a Product Lifecycle Management system (PLM).

STEP 3: DATA PREPARATION

The data is extracted from the relevant data sources using an ETL process (extraction, transformation, loading), compressed by means of transformation, corrected, harmonized, and then loaded into the data warehouse.

This process is carried out regularly in order to keep the data in the data warehouse up-to-date. To allow comparative analyses of a large data quantity with the relevant history, the data can be stored for long periods. In other words, a location with redundant information is deliberately created with the aim of ensuring that the data analytics software has access to all important information at all times. Transformation and compression reduce the data quantities considerably, allowing rapid comparison and user-friendly analysis.

A patented compression and transformation algorithm was developed for data preparation. This allows the engineer to store large engine maps without taking up too much memory space and swiftly compare them [6].

STEP 4: MODELING

The template approach described above was chosen for modeling. It is comparable with the "personalized recommendations of items represented within a database" patented by Amazon [7].

STEPS 5 AND 6: EVALUATION AND ROLL-OUT

The user-friendly Compare Pro software was developed to make data analysis as easy as possible. This is provided within AVL CRETA.

On the basis of the CRISP-DM process, it is now possible to carry out a cross-comparison of more than 2,000 data sets within an extremely short time.

MORE EFFICIENCY AND HIGHER DATA QUALITY

When data analytics is used for variant calibration [7], the effort pays off after an extremely short time. The definition of the attributes is a one-time effort. Afterwards, the structure can be used for all projects. Only relatively small adjustments of the attribute structure are necessary over the entire runtime of the project.

The tool presented here allows the individual calibration engineer to search for specific information on the parameters for which he is responsible. In addi-

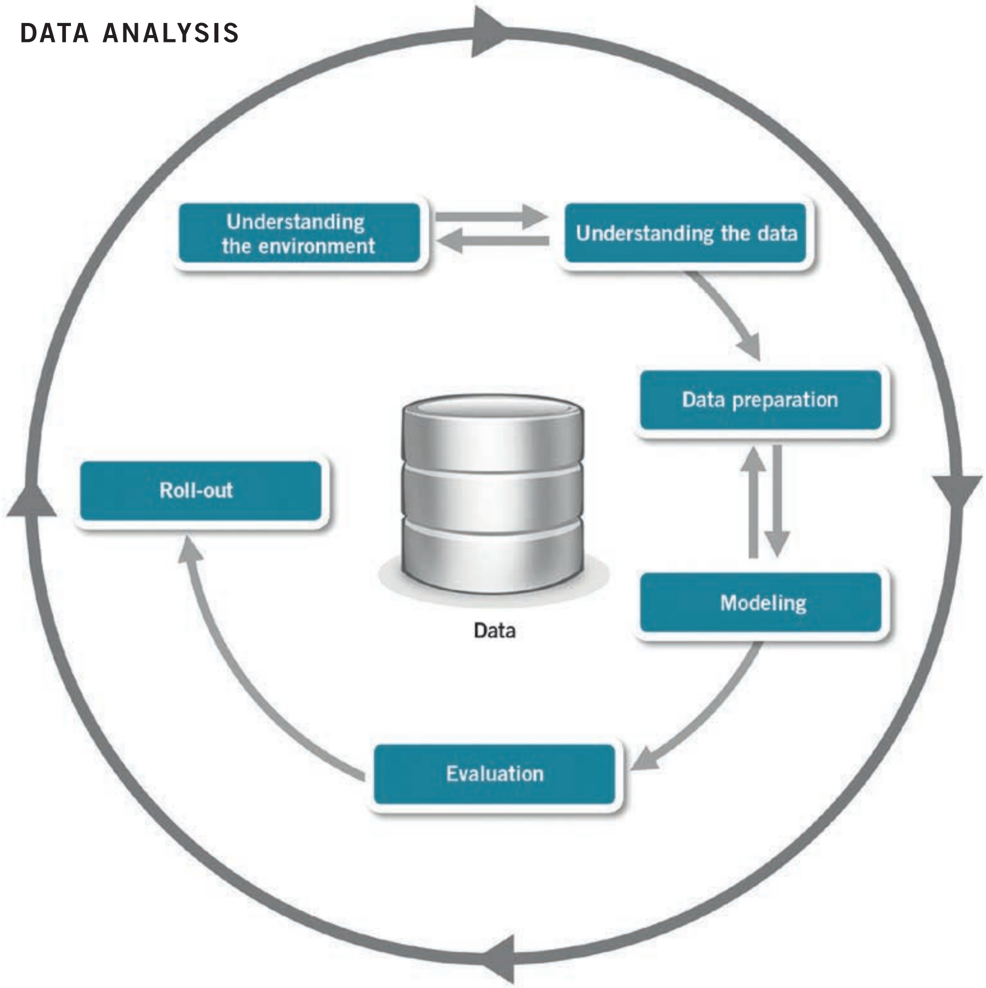


FIGURE 6 Steps in the CRISP-DM process model (diagram based on [5]) (© AVL)

tion, he can check his data sets for individual outliers before transferring them. Data analysis also brings consistency and success to the variant calibration process as a whole. The templates allow the engineer to check regularly whether all variants comply with the templates validated by the expert. Unusual calibrations are immediately obvious.

To derive standard original calibrations for future projects, the templates can be managed in a library. The calibration teams can thus start their work on the basis of a mature data status and achieve good results at an early stage of the development process.

After completion of a project, data analysis can make valuable contributions to the evaluation of the calibration effort. How did the engineers arrive at the final calibration values? Which labels were not changed at all or only once? Which projects changed unusually frequently in the project period and why? These questions make valuable contributions to the planning of future projects.

CONCLUSIONS

The rapid data growth in calibration projects makes new methods and tools urgently necessary. Data analytics is a promising approach here. However, only a very small number of companies make use of the knowledge contained in their calibration data. This newly developed tool allows companies to use data analytics for smart variant calibration with the aim of improving the efficiency and quality of their calibration projects.

BIBLIOGRAPHY

[1] Keuth, N.; Kordon, M.; Wurzenberger, J.; Vitale, G.: Paradigmenwechsel in der Kalibrierung: Innovative Methoden zur Erhöhung der Qualität der xCU Kalibrierung. 5. Intern. Symposium für Entwicklungsmethodik, 2013
[2] Paulweber, M.; Lebert, K.: Mess- und Prüfstandstechnik. Wiesbaden: Springer Vieweg, 2014, S. 394-397
[3] Dobes, T.; Jansen, H.; Kaserer, T.; Klumaier, K.; Riener, M.: Moderne Entwicklungsplattform für die Kalibrierung von Fahrzeug-Antrieben. 6. Intern. Symposium für Entwicklungsmethodik, 2015
[4] Rathfelder, M.; Hsu, H.; Brandau, T.; Storfer, G.: Applikationsdaten-Management für Porsche-Fahrwerke. In: ATZ 118 (2016), Nr. 6, S. 16-20
[5] Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R.: CRISP-DM 1.0. Step-by-step data mining guide, 2000. Online: http://www.the-modeling-agency.com/crisp-dm.pdf, aufgerufen am 20.02.2018
[6] AVL List: EP2859415B1: Verfahren zur Verarbeitung von Daten, 15.04.2015
[7] „Kunden, die dieses Buch gekauft haben, haben auch dieses Buch gekauft." Vergleiche dazu: Amazon.com, Inc: US7113917B2: Personalized recommendations of items represented within a database, 13.09.2001

THIS ARTICLE IS AVAILABLE IN THE E-MAGAZINE UNDER:
www.emag.springerprofessional.de/atz