Utilization of AI-Driven Solutions for Rubber Compound Formulation: A Practical Approach

Hans-Joachim Graf, H-JG Consulting
Christoph Hartwich, IT Consultant

Abstract:

The integration of artificial intelligence (AI) into various industrial sectors has spurred remarkable progress in the field of manufacturing process development. This concept has been extended to polymer compound formulations, where AI techniques have been used to optimize rubber formulations (**1-10**) as well. Furthermore, it is conceivable that mixtures of other polymers, such as TPEs, thermoplastic materials, or polyurethane compounds, could be similarly treated. Advantages of employing AI-driven tools include reduced development time, enhanced efficiency, and increased accuracy.

However, certain formidable challenges exist around AI-based rubber formulation development (**11**). One identified consideration is that these tools are useful only when conventionally organized datasets, cultivated over years, through rigorous experimental design techniques, are used. AI analysis is only as good as the dataset it is analyzing.

Besides the challenges associated with using unstructured databases, additional difficulties arise from the fact that many mixtures address multiple and highly variable desired solutions. Finally, specific to rubber development, it is also understood there is some degree of measurement error compared to the exact values that are obtained in other fields.

Due to these inherent challenges, at present, the use of AI in rubber compound development is generally restricted to small, highly specific databases and can only be used by experienced, knowledgeable technicians.

To address these challenges, this article presents two possible solutions:

1. Certain pragmatic methodologies for handling data under non-ideal conditions is described. They aim to handle the absence of systematic organization and offer practical insights into navigating the complexities of unstructured data.
2. An AI tool, specifically tailored for accurately formulating mixtures from data characterized by diverse testing conditions, inherent errors and limited structural organization, is also described. By leveraging advanced AI algorithms, this tool can mitigate the negative impact of data variability and incompleteness, thereby enhancing the accuracy of mixture formulation prediction.

Through its comprehensive exploration of challenges and innovative solutions, this article contributes to the evolving discourse on the practical implementation of AI tools in rubber compound development.

Introduction

The importance of effective mixing formulation development within the rubber industry is

undeniable. Rubber compound formulations have an influence not only on product quality and functionality but also on manufacturing efficiency. As mentioned, AI technologies (a sketch of a forward propagation AI is depicted in **Fig. 1**, have already made their way into the manufacturing processing sector. As developers become convinced of the benefits of AI-driven compound development as a powerful assistant, its integration within the thermoplastic and rubber compounding fields will most likely advance.

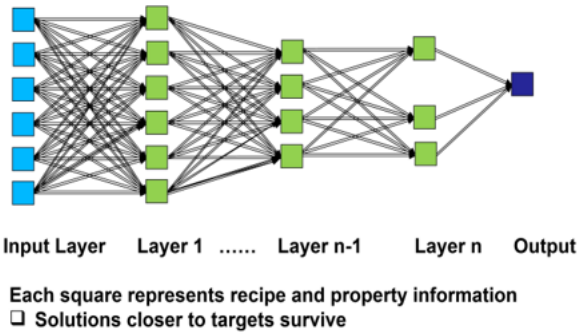

Fig. 1: Principle drawing of a multi-layer forward propagation AI

Along with this burgeoning development however, challenges related to using AI tools on unstructured datasets have arisen. In the late 1990's, the author had the opportunity to test a - let us say – first generation AI tool. Despite conducting tests with a consistent dataset of several hundred mixtures based on only one EPDM polymer, significant deviations in predicting new mixture properties arose. The primary reason for these failures lied in the lack of transparency or back traceability. The deficiency between the data and confirmation experiments was extremely large, indicating the presence of unknown faulty data. Errors within the data, for instance, measurement errors or even typing errors, could be neither identified nor rectified.

Since the introduction of experimental design techniques that construct numerically limited datasets through variations of a few ingredients, attempts have been made to push these boundaries and overcome dataset challenges (**12-14**). However, the majority within the rubber industry still tend to utilize historically accrued files for development and thus, the chance for significant AI analytic errors still remains.

AI Analysis Challenges in Rubber Compound Development

Implications for AI Precision – Unstructured Databases

Many companies currently use and populate databases that have limited structural organization. For example, these databases may contain non-comparable values, i.e., those that have been measured according to different testing protocols like ISO/ASTM. Values may also have come from non-standard customer specifications. They may even contain missing property values.

Unfortunately, within the rubber industry, these types of databases are more prevalent than the structured, logical type. The resultant dataset values therefore, cannot be considered comparable and as such, cannot be analyzed using AI tools with any accuracy. This represents a significant impediment for the rubber industry in the use of AI-driven tools for development.

Implications for AI Precision – Incompatibilities

In the realm of AI and cloud computing, it is widely acknowledged that AI thrives on copious amounts of data, and, with the assumption "the more data, the better," precision in prediction is

typically assured. While this assertion resonates in contemporary AI applications, the landscape differs significantly in the rubber industry.

This industry performs redundant experiments in many cases, even when employing DoE techniques. Unfortunately, the conventional DoE approach becomes impractical due to the industry's complexity, where numerous compound ingredients cannot be held constant while varying three to four ingredients. The resulting influx of experiments poses challenges in both laboratory execution and data analysis complexity.

As a result, more and more data from experiments accumulates, which in turn requires a meticulous search-and-find approach for a few target property analyses. Additional difficulties arise from the fact that many mixtures address multiple and highly variable solutions. For example, certain applications may require different base polymers or crosslinking systems. For instance, sulfur-based systems may work well for one application compared to a peroxide-based one. Or, an HNBR compound may work better for one application compared to EPDM. Or even, two entirely different base compounds/crosslinking systems may suitably work for one application. The database necessarily ends up containing data on multiple polymers or crosslinking systems. Additionally, both ingredients and properties can be unevenly distributed. This “incompatibility” between polymers or cross-linking systems can lead to unmanageable second-order effects.

Therefore, predictive mixtures, because they reference completely different base polymers/crosslinking systems, etc., are not directly discernible in extensive, pre-existing datasets. There is no way to easily separate such datasets. This again, poses a challenge for AI tools tasked with generating viable solutions.

Implications for AI Precision – Data with Unpredictable Errors

In rubber development, it is understood there is some degree of divergent error compared to the exact, pristine values observed in other fields in the computing domain, such as spell checkers or face recognition. Specifically, divergent error can be represented by the equation:

(1) $y = x + \mu$ (with x = true value, μ = error)

This transcends the traditional definition of error. Unlike datasets in other fields, errors in rubber testing are not only non-normally distributed, but can also exhibit diverse sudden onsets, among other possible distributions. These induced errors that occur through, for example, the change of procedures (mixing process, test protocols), testing equipment, material storage degradation, or even operator influence can find their way into the data over time. Additionally, measurement errors can be introduced during test sample preparation, for example, changing storage/preconditioning conditions, testing on different equipment, or even, movement of the testing facility to other locations.

As an illustration, consider the crucial parameter of tensile at break [TB] and the intricate network of factors that influence its true value.

Variables such as the molecular weight (MW) of the polymer, energy uptake during mixing, filler distribution and dispersion, variations in the weight of crosslinkers and accelerators, for example, all contribute to the nuanced nature of TB. This is over and above the aforementioned environmental influences. Considering all these factors will help provide a comprehensive understanding of the complex error distribution in rubber development.

Unfortunately, the cumulative impact of diverse sources of error cannot be accurately mathematically modeled. This reality sheds light on the challenges posed and their implications for AI prediction precision. It makes certain machine learning [ML] tasks particularly challenging within this domain when using forward propagation AI tools.

Practical Solutions When Working With AI

Dealing with Historically Accumulated Mixture Databases

The following solution is proposed when analyzing massive amounts of disparate information in one large file. Contrary to conventional wisdom, storing all this data in a single file is ill-advised. Going forward, such data sets should be divided, possibly using the following separation criteria:

- Polymers (based on their compatibility),
- Vulcanization systems (based on sulfur, peroxide, metal oxide, resin or other chemicals).

This separation will necessarily result in a large number of smaller datasets. Consequently, **if AI-driven analysis is to be successful within the scope of this application, it needs to be prepared to handle smaller amounts of data**. However, the downside to using smaller datasets is that there might be insufficient data available for a potential solution or, a prediction may occur that lacks statistical confidence.

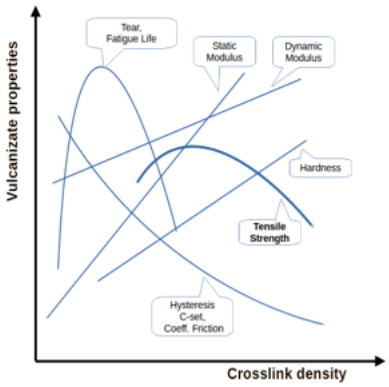

Fig. 2: Physical properties versus crosslink density - sketch published by Coran

Methodology to Address Non-Ideal Data Conditions

One pragmatic solution to address dataset challenges involves detecting and eliminating faulty data using a correlation analysis between ingredients and properties, as elucidated in the renowned diagram by Coran and Studebaker (**15 - 17**) (**Fig. 2**). This visual representation effectively captures the correlation of crosslink density to physical properties in vulcanized rubber compounds. It describes a compound at constant carbon black and oil loading where the crosslinking agent (sulfur) is the only variable. While it is widely acknowledged that additional components, such as carbon black, oil, polymer and crosslinker, can significantly influence TB and other properties, these factors are expected to adhere to the correlation demonstrated in Coran's diagram.

Coran's diagram needs only to be converted into a multidimensional format to be useful for more in-depth data analysis, as illustrated in **Fig. 3.** A Cartesian diagram, with properties plotted against crosslink density, are constructed for each level of carbon black and oil loading, forming perpendicular representations. In this illustration, utilizing datasets from published sources (**18, 19**), one can effectively showcase the influence of variations in carbon black and oil loading, as well as their types (represented on the z-axis), on the tensile at break (TB on the y-axis).

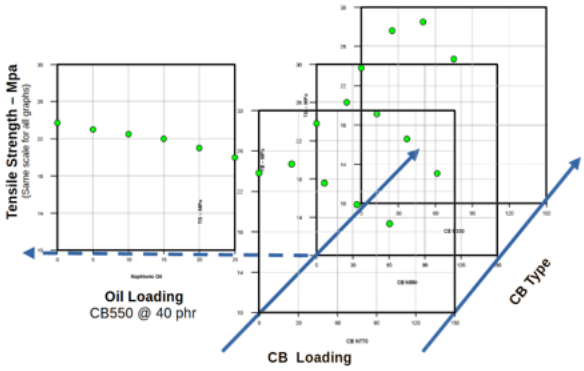

Fig. 3: TB versus oil loading at one datapoint – CB N550 at 40 phr
TB versus CB type and loading

As another example, the impact of sulfur as the crosslinking agent suggests that the sulfur bridge structure undergoes changes with varying sulfur amounts, influencing the tensile at break relative to crosslink density (sulfur on the x-axis, or alternatively MH-ML on the x-axis) (**Fig. 4**). These curves were generated using a sulfur-cured Natural Rubber (NR) compound accelerated with N-Cyclohexylbenzothiazol-2-sulphenamide (CBS), maintaining a constant sulfur-to-CBS ratio, at two different levels of sulfur (**20**). This multidimensional approach provides a comprehensive view of the interrelationships between these crucial factors.

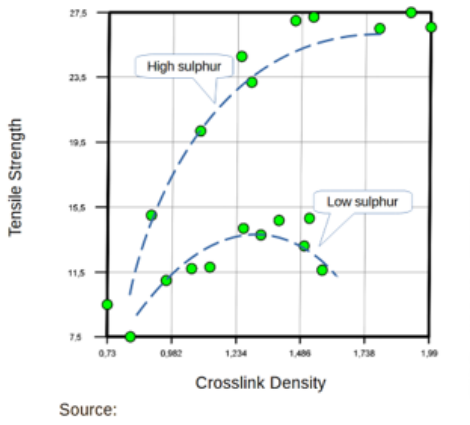

Fig. 4: Tensile strength versus Crosslink Density at two different Sulfur levels

Dealing with Data Containing Unpredictable Errors

It is challenging to make definitive judgments about errors in the data, especially when dealing with larger datasets. There are two ways of tackling this:

- When confronted with repetitive data, valuable insights into measurement errors can be gained. Traditionally, the control compound acts as a benchmark, as is used in Design of Experiments (DoE) in combination with iterative processes. Thus, it is advisable to incorporate a control compound as a standard procedure, even in trial-and-error experimentation. Adhering to good manufacturing practice (GMP), it is recommended to produce and test the control compound three to four times a year, for example. This precaution is essential due to the inherent unpredictability of results stemming from the aforementioned measurement errors and ensures no erroneous conclusions are drawn.
- Quality dataset assessment can be facilitated again, using correlations between ingredients or among properties themselves. The user can generate 2D graphs that illustrate relationships between ingredients and/or properties, along with regression correlation

coefficients. Additionally, users can employ a 3D graph for the analysis of relationships involving three variables, such as two ingredients and a specific property, for example.

- The generation of diagrams depicting property versus crosslinker concentration, typically used as a measure of crosslink density, while keeping other ingredient concentrations constant, is a crucial step. If the resulting diagram, using the same x-y axis as in Coran's work, demonstrates a similar curvature or trend with minimal scatter around that curve, the condition proposed by Coran's scheme is fulfilled. In such cases, the error can be estimated through visual data inspection or with regression and its correlation coefficient.

As an example, using a dataset of 23 compounds, a 3D graph can be generated, placing F_{max}-F_{min} values on the x-axis for use as a synonym for crosslink density (**21, 22**). Sulfur levels of 0.6 phr, 1.25 phr, and 2.0 phr are plotted on the y-axis, and tensile strength on the z-axis (**Fig. 5**). The graph adequately follows the trends of Fig. 4, albeit with slightly more scattering around the trend curve.

To increase measurement confidence, standard deviations and errors should be assessed to decide whether datasets should be retained or excluded. To aid in differentiation between a measurement value and the "true" value, it is sensible to exclusively utilize data either from certified laboratories or from data that is trusted by the user, with minimized measurement error.

Formulations for rubber compounds or other polymer compositions, such as TPEs, thermoplastic materials, and polyurethane compounds, can be used as long as they are normalized formulations, as customary in the rubber industry.

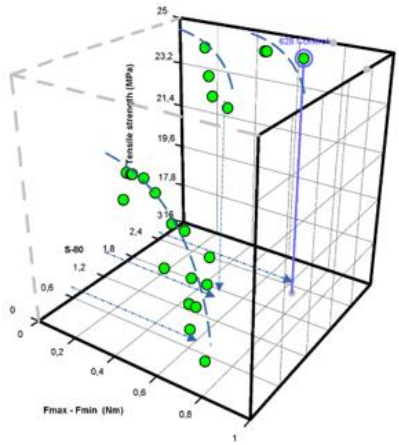

Fig. 5: Tensile Strength versus Fmax-Fmin and Sulfur at 3 different levels (0,6 phr, 1,25 phr and 2,0 phr)

Optimizing AI Integration in Rubber Industry Data Management

Thus, when working with existing rubber data information, from an AI perspective, it is imperative that smaller data amounts stored in discrete repositories be used for analysis. In general, compound data sets should adhere to a standardized structure. Foundational physical properties should be consistently measured for each compound, complemented by customer-specific attributes. Similarly, if Trial-and-Error, One Step at A Time or DoE techniques are employed, adherence to standardized protocols is imperative.

While the standard structure involves conducting a larger number of tests per compound in the development phase, the subsequent benefits of utilizing AI tools are noteworthy. These advantages encompass the prevention of redundant experiments through the prediction of an "in-specification" compound, which should always be verified in a confirmation experiment. In the event of an unacceptable deviation between the predicted compound and confirmation, it serves as a good starting point for further refinement.

To facilitate seamless data integration, standardization of ingredient names, property names, and procedural and employee nomenclature should be implemented. Such standardization ensures

compatibility, allowing for the effortless merging of compound data files with minimal effort and complication. In essence, the rubber industry's journey towards AI integration is underlined by the need for strategic data structuring and standardized practices to unlock the full potential of advanced analytics.

AI Development Tool for Rubber Compounds - A Practical Approach

A software tool based on a forward propagation AI called "GrafCompounder" (**Fig. 6**) seems particularly suited to address many of the above concerns (**23**).

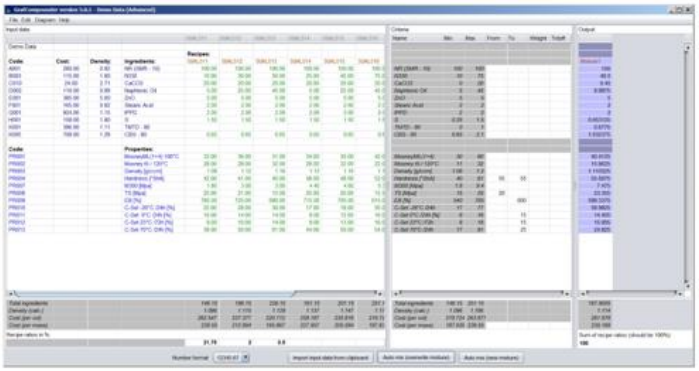

Fig 6: Screen shot of GrafCompounder AI software tool

The aforementioned prediction failures prompted the development of this AI tool (**24, 25**). It is grounded in a highly pragmatic approach. It does not possess any so-called "secret" domain knowledge. Most importantly, it does not endeavor to derive equations, analytical or regression, from the assigned data. Instead, the tool treats each formulation as a "data point," consisting of a list of ingredients with corresponding quantity indications and a list of properties with their measured values. Each data point can be described as an n^{th}-dimensional vector, where n represents the number of ingredients and properties.

At its core, this tool employs a specialized AI technique, a highly specific genetic algorithm, that converges on a solution through multiple intermediate steps (generations). The process operates automatically and interconnects data points through simple interpolation in small steps. It determines which data points / vectors should be mixed and which factors should be used to best fulfill the user-defined target mixture properties. It addresses both the selection of the most suitable data points and, if necessary, the determination of optimal compromises among all requirements. To achieve this, the process internally employs a computation called the "fitness function," which it generates from the given criteria.

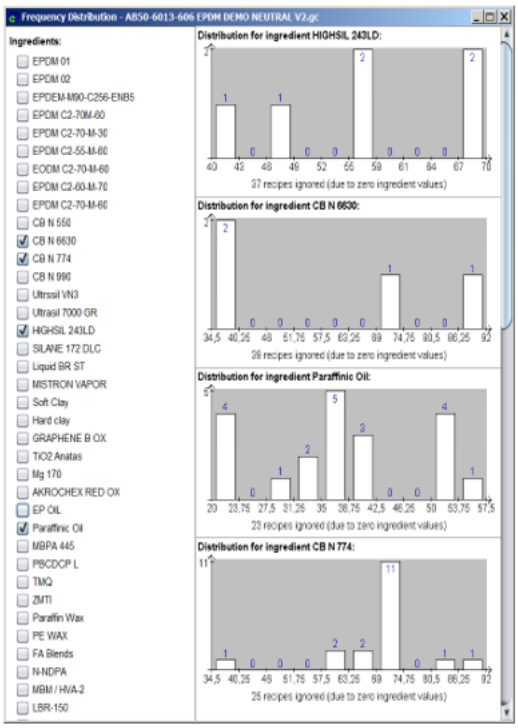

Fig.7: Frequency distribution bar diagram of selected ingredients in small compound data set

Its ultimate advantage is that it can work with existing datasets without any modifications. Limitations arising from the dataset's structure must be accepted or modified by the user. Several tools and features are available in the software for this purpose, for example, a frequency bar diagram (**Fig. 7**):

- The frequency distribution of each mixture component and / or each test value can be displayed and evaluated. This tool allows the user to identify empty data spaces that should be filled with additional compound data.
- Compound data points with data missing in essential places, like missing values for properties specified as criteria, are automatically excluded from calculations.
- Datasets from other files can be added and integrated with the merger function.
- The composition of the predicted mixture is comprehensively presented, ensuring backward traceability to the original formulations utilized during the prediction process.

Ultimately, this tool can counteract some of the disadvantages of dealing with non-structured / disparate databases.

Considerations and Outlook

The integration of AI into mixing formulation development brings about practical advantages, including:

- **Efficiency:** A systematic development approach that minimizes experimental efforts, thereby expediting the development process.
- **Accuracy:** Utilizing AI-driven tools enables targeted variations in criteria to focus on specific material properties.
- **Optimization:** It has the capability to improve existing formulations and tailor them to specific requirements.
- **Sustainability:** The reduction of waste is achieved through decreased experimentation efforts. This enables the analysis of numerous formulations, with only a few selected for actual confirmation experiments.

However, the absence of systematically constructed historical datasets, material incompatibilities, and the possibility of inherent measurement errors poses a challenge, requiring a severely structured, pragmatic approach in the application of AI-driven tools.

Consequently, the use of AI tools at present requires technological knowledge and judgment not only in selecting and compiling data for calculations but also, based on the user's experience and knowledge, their ongoing assessment of whether the results make sense. Consequently, existing AI tools for mixture development should currently be classified as **assistance systems** and utilized only by expert rubber technologists.

The GrafCompounder AI tool is poised to work well for processing historically accumulated datasets by the general user, particularly if assessment criteria such as data reliability can be established for AI processing. The tool has been designed with the goal of contributing significantly to the efficient, accurate, and sustainable advancement of rubber compound formulation development.

Conclusion

Given the inherent challenges associated with current rubber formulators' existing datasets, the employment of general AI-based analytical tools can only be used in rigorously controlled instances and cannot guarantee statistically significant results. However, because of its promising inherent

ability to analyze mass amounts of information, it is imperative that its use be advanced within the rubber industry to avoid the loss or neglect of valuable historical data.

The AI-driven development tool "GrafCompounder" represents a significant advancement in polymer formulation development. Unlike traditional trial-and-error approaches, this tool allows developers to integrate their expertise with AI-based calculations. The incorporation of this tool into rubber compound development procedures, including data management, enhances data quality and provides more accurate predictions for rubber compounds, accelerating development and encouraging the creative exploration of new possibilities.

References

1. K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, John Wiley & Sons, Inc., 2001
2. A. C. C. Coolen, A Beginner's Guide to Mathematics of Neural Networks in: Concepts for Neural Networks – A Survey, L. J. Landau, J. G. Taylor (Ed.), Springer, Berlin, 1998, 13
3. Schwartz, G. A. Prediction of rheometric properties of compounds by using artificial neural network. *Rubber Chem. Technol.* 2001, *74*, 116-123
4. Y. Nakajima, Inventor; US Patent 06411965B1, June 25, 2002
5. R. C. Rowe, E. A. Colbourn, Neural Computing in Product Formulation. Chem. Educator 2003, 8, 1-8
6. Y. Shen, K. Chandrashekhara, W.F. Breig, L.R. Oliver; Neural Network Based Constitutive Model for Rubber Material. *Rubber Chem. Technol.* 2004, 77, 257-277
7. V. Vijayabaskar, R. Gupta, P. P. Chakrabarti, A. K. Bhowmick, Prediction of properties of rubber by using artificial neural networks. J. Applied Polymer Sci. Vol 100 (3), 2006
8. U. Zeyneb, A. Kiraz, Artificial Intelligence based prediction models for rubber compounds. J. Polymer Eng. 43 (2022), 2, 113-124
9. Y. Shen, K. Chandrashekhara, W.F. Breig, L.R. Oliver; Neural Network Based Constitutive Model for Rubber Material. Rubber Chem. Technol. 77 (2004), 257-277
10. C. Drahl, J. Kemsley, AI for Materials discovery, ACS Report; C&EN 02 (2021)
11. J. Huang, J. Liew, A. Ademiloye, K. M. Liew, Artificial intelligence in materials modeling and design. *Archives of Computational Methods in Engineering*, (2020) *28*(5), 3399–3413. https://doi.org/10.1007/s11831-020-09506-1
12. EP 0865 890 A1, Method of Designing Multicomponent Material, Optimization and Storage Medium on which Multicomponent Material Optimization Analysis Program is Recorded, Applicant: Bridgestone Corporation, (1998)
13. US 7541122B2, Empirical Design of Experiments Using Neuronal Network Models, Assignee: Honeywell International Inc. (1996)
14. H-J. Graf, Methodical Recipe Development: Increases Technological and Economical Advantages [GE], Conference Proceeding at IRC 2009, June 29. - July 2., Nuremberg, Germany, (2009)
15. P.J. Flory, N. Rabjohn, M.C. Schaffer, Dependence of Tensile strength of Vulcanized Rubber on the Degree of Crosslinking, Rubber Chem. Technol. 23 (1950) 27-43
16. M.L. Studebaker, Effect of Curing systems on Selected Properties of Natural Rubber Vulcanizates, Rubber Chem. Technol. 39 (1966) 1359-1381
17. Coran, A.Y., in "Science and Technology of Rubber," F.R. Eirich, Editor., Academic Press, New York, (1978) Ch. 7, p. 292.
18. R.F. Ohm, Editor, The Vanderbilt Handbook, Thirteens Edition (1990)

19. R. Kirschbaum, F. Jones, Profile of Cabot Carbon Blacks in Natural Rubber, Technical Report RG 133
20. L. Gonzalez, A. Rodriguez, J. L. Valentin, A. Marcos-Fernandez, P. Posadas, Conventional and Efficient Crosslinking of Natural Rubber - Effect of Heterogeneities on the Physical Properties, Kautsch. Gummi Kunstst. December (2005), 638-643
21. S. Lee, H. Pawlowski, A. Y. Coran, Method for Estimating the Chemical Crosslink Densities of Cured Natural Rubber and Styrene-Butadiene Rubber, Rubber Chem. Technol. 67 (1994) 854-864
22. M. Saewe, Investigation of Sulfur Vulcanization of Polybutadienes with Benzthiazole-sulfenamides, Thesis (GE) University of Hannover, Germany (2007)
23. H-J. Graf, A New Tool for a Systematic Development and Improvement of Compounds [ENG], Paper presented at Rubber Division of ACS, 183 Spring Meeting, Akron, OH, April 24. (2013), USA
24. H-J. Graf, C. Sirisinha, P. Pangamol, Simulation of a statistical experimental design with a new software tool, Rubber Fiber Plastics 04 (2015) (Vol 10), 266-276
25. H-J. Graf, Where are we with Artificial Intelligence & Machine Learning in Rubber Development? Poster Presented at RubberCon23, 10-11 May, Edinburgh, Scotland (2023)