Semi-automatic Evaluation of the Degree of Bitumen Coverage on Bitumen-Coated Aggregates

Riccardo Lamperti, Claudio Lantieri, Cesare Sangiorgi, Gabriele Bitelli and Andrea Simone

Abstract EN 12697-11 is the standard providing test methods for evaluating affinity between aggregate and bitumen and its influence on the susceptibility of the mixture to stripping. Among the methods, the rolling bottle test has a number of advantages in terms of rapidity, low costs, and suitability for routine testing. However, since affinity is assessed by visual registration of two independent operators, results may be altered by a large amount being inevitably subjective. The authors suggest a semi-automatic procedure to overcome potential limits and shortcomings of the method and obtain accurate results. Different mixtures were analyzed and compared, using common natural and recycled aggregates and a 70/100 pen bitumen. The procedure was successfully validated with a manual pixel inspection and confusion matrixes were created. The results showed that the procedure lead to a more reliable registration compared to the standard method and it is suitable, with different accuracies, for both light and dark aggregates.

Keywords Rolling bottle test · Adhesion · Bitumen · Confusion matrix

1 Introduction

Asphalt mixture is a complex and heterogeneous material that includes aggregates, asphalt binder and air voids. Its overall mechanical response is primarily governed by the asphalt binder and by the stone-on-stone contacts between aggregates (Dondi et al. 2012; Vignali et al. 2014). Loss of adhesion between bitumen and aggregates and loss of cohesion within the mixture, in the presence of water, are referred to as moisture damage and represent one of the main causes of distress in asphalt pavements. The adhesion quality (i.e. affinity) between binder and aggregate is

R. Lamperti (✉) · C. Lantieri · C. Sangiorgi · G. Bitelli · A. Simone
Department of Civil, Chemical, Environmental, and Materials Engineering—DICAM, University of Bologna, Bologna, Italy
e-mail: riccardo.lamperti2@unibo.it

F. Canestrari and M.N. Partl (eds.), *8th RILEM International Symposium on Testing and Characterization of Sustainable and Innovative Bituminous Materials*, RILEM Bookseries 11, DOI 10.1007/978-94-017-7342-3_2

essential for the stability, quality and endurance of the road under climate and traffic. Although not all damage is caused directly by moisture, its presence increases the extent and severity of already existing distresses like cracking, potholes, and rutting (Miller et al. 2003). The presence of moisture results in a degradation of the mechanical properties of the asphalt mixture, i.e. loss of stiffness and mechanical strength, which ultimately leads to the failure of the road structure (Grenfell et al. 2014). Many researchers have recognized that the replacement of bitumen film from the aggregate surface by water, referred to as stripping, is linked to interfacial tension relations of these materials (Hefer 2004; Bhasin 2006).

The standardized approach to quantify the affinity between aggregate and bitumen is the rolling bottle test (EN 12697-11). It consists in placing a mix of bitumen and aggregates in a bottle filled with de-ionized water and then placing it in a rolling machine in order to subject the material to a mechanical stirring action in the presence of water. After defined time steps, normally 6 and 24 h, two independent operators visually estimate the residual degree of bitumen coverage of the particles. Despite being a rapid, simple and low costs test, results may be altered by a large amount being the determination of the bitumen coverage degree inevitably subjective. The main factors that may influence the estimation are linked to the skills of the operators, the light conditions and the color of the aggregates. Dark or grey aggregates like basalt or blast furnace slag may be confused with bitumen. Different attempts were made in order to improve the determination of the degree of bitumen coverage of aggregates after the rolling bottle test is performed. Mulsow (2012), based on the observation that the micro-roughness of the surface of the aggregate is significantly higher than bitumen, studied the adhesion with a multi-directional reflectance measurement. Grönninger et al. (2010) and Källén et al. (2012) used supervised classification and advanced segmentation methods of RGB images respectively. Despite these efforts, a standard and validated procedure has not been found yet.

2 Experimental Work

The aim of this paper is to propose an alternative assessment method in order to overcome the potential limits of the visual estimation procedure of the rolling bottle test. A series of rolling bottle tests was performed. Three independent skilled operators made visual observation of the samples after 6 and 24 h. Furthermore, at each time step, digital pictures of the samples were taken and processed with a software to obtain a second estimation. This estimation is based on RGB images that are subjected through a series of filters, first deleting the background, and then isolating the only bitumen pixels. The affinity estimation of blends is made by computing the areas of the bitumen and of the aggregates on the images.

Results of both estimations and a comparison between them is presented. The procedure was validated by checking the accuracy of the software measurements on a 51 × 51 grid of points on the images, for each different adopted aggregates. This lead to create confusion matrices to interpret the accuracy of the software estimation.

2.1 Materials and Equipment

Four different aggregates were tested: porphyry, limestone, basalt and blast furnace slag (Fig. 1). They were selected in order to cover a broad range of colors. All aggregates were sieved and washed to obtain a 8–11 mm fraction, according to standard. The adopted base bitumen was a 70/100 pen. Three bottles were tested for each mix for a total of 24 observations.

The materials were heated at 160 °C. All the procedures described in EN 12697-11 for mixing and testing were followed. Rolling speed was 60 rpm and the tests were performed at a temperature of 20 ± 1 °C. Visual observation were made after 6 and 24 h and digital images were taken accordingly.

2.2 Research Approach

The use of a computer aided analysis technique is suggested and is based on digital picturing of the bitumen-aggregate sample after the rolling phase and on the classification of characteristic color areas by means of a commercial software (Grönniger et al. 2010). In this regard a specific photographic set was developed (Fig. 1a). The tested blends were put into a plate filled with de-ionized water. The plate had a green background and was irradiated with 2 lamps with an angle of incidence of the light beam of 45° in order to avoid reflections. Picture were taken vertically with a 10 MP camera, with ISO 100, from a distance of 30 cm. Pictures were further processed with a public domain Java-based imaging program. Among the different color spaces available, the YUV was chosen. The YUV color space separate RGB into luminance and chrominance information (Ford et al. 1998) in terms of one luma (Y) and two chrominance (UV) components. Luma stands for brightness, or lightness, while U and V provide color information and are "color difference" signals of blue minus luma (B-Y) and red minus luma (R-Y). YUV was

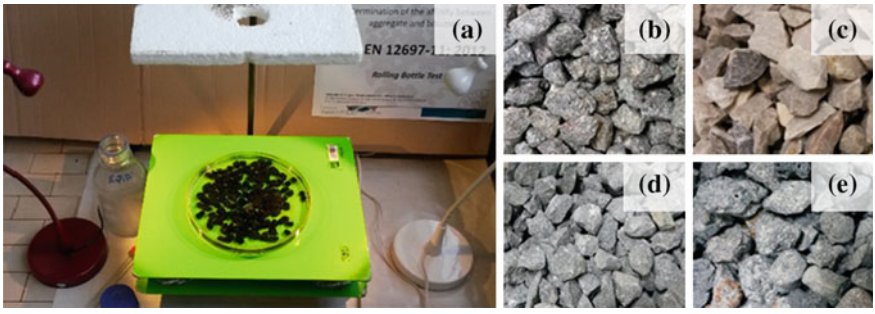

Fig. 1 **a** Set for digital picturing of tested samples; **b** Porphyry; **c** Limestone; **d** Basalt; **e** Blast furnace slag samples

Fig. 2 Image filtering process. **a** Original image; **b** Non-classified image; **c** Classified image

chosen because of its ability to decouple the luminance and color information where the image can be processed with no effect on the other color components (Ibraheem et al. 2012).

This peculiarity was confirmed after testing the other available color spaces (i.e. HSB, RGB and CIE Lab) which gave worse results in terms of quality of the image classifications.

The adopted software allows defining a specific threshold for each of the Y, U & V component that compose the image. Therefore, acting on these ranges the image was filtered, leaving out the background. Once the image contained only aggregates and bitumen pixels, referred to as *non-classified image* (Fig. 2b), the Y, U & V components ranges were further reduced in order to get only the bitumen pixels as shown in the *classified image* of Fig. 2c. At each step, the program allowed measuring the selection areas, i.e. the sum of pixels within the specified YUV ranges. The bitumen plus aggregates and the only bitumen areas were automatically computed, enabling the calculation of the percentage of the bitumen coverage of the aggregates with Eq. (1):

$$Bitumen\ coverage = \frac{A_{bitumen}}{A_{bitumen+aggregates}} \times 100\ [\%]. \tag{1}$$

3 Results

Table 1 shows the YUV ranges found for the image identification of the studied bituminous materials. Y, U & V can vary from 0 to 255. Lower or upper limits vary with different exposure conditions, i.e. with natural light, artificial light only or both. In order to reproduce the analysis of the same materials, an operator could apply these YUV sets and easily obtain the non-classified and the classified images. For example, for the identification of U parameter of the Limestone in the classified image, the lower limit is chosen between 80 and 100 and the upper limit is 255.

Table 1 YUV ranges for materials' recognition

		Limestone (L)	Porphyry (P)	Basalt (B)	Blast furnace slag (S)
Non - classified image	Y	0–195	0–165	0–165	0–195
	U	(80 ÷ 100) − 55	(80 ÷ 90) − 255	(90 ÷ 110) − 255	100–255
	V	(110 ÷ 120) − 255	(115 ÷ 120) − 255	(105 ÷ 120) − 255	120–255
Classified image	Y	0 − (55 ÷ 85)	0 − (25 ÷ 35)	0 − (25 ÷ 45)	120–255
	U	(80 ÷ 100) − 255	(80 ÷ 90) − 255	(90 ÷ 110) − 255	100–255
	V	(110 ÷ 120) − 255	(110 ÷ 120) − 130	(105 ÷ 120) − 255	120–128

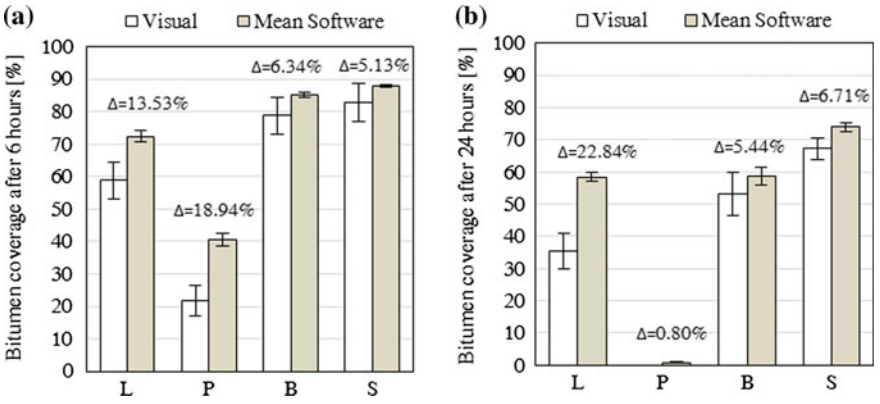

Fig. 3 Results of rolling bottle tests after 6 h (**a**) and 24 h testing (**b**)

Figure 3 shows the results of the tests. The white columns represent the visual estimation carried out by three independent skilled operators, while the dark columns represent the software semi-automatic classification. Results are shown in terms of mean (%) and standard deviation of all the observations, i.e. three for the software and nine for the visual estimations (three bottles and three operators).

Considering the software recognition, the limestone exhibited a percentage of coverage at 6 h of 72.4 ± 1.6 and of 58.4 ± 1.4 after 24 h. The porphyry was the more prone to stripping (40.6 ± 2.0 after 6 h and 0.8 ± 0.2 after 24 h). The basalt showed a higher bitumen coverage after 6 h 85.2 ± 0.8 and reached a coverage comparable to that of limestone after 24 h (58.8 ± 2.6). The blast furnace slag gave the best performance compared to the other aggregates (87.91 ± 0.49 after 6 h and 73.93 ± 1.48 after 24 h).

Furthermore Fig. 3 allows for a comparison between the software and the visual registrations. Regardless of the accuracy, light colored aggregates (L & P) exhibited the higher gap between the visual estimation and the software recognition, with a maximum difference of 22.84 % for the limestone after 24 h. On the other hand, differences of recognition for dark aggregates like basalt and blast furnace slag

amounted to 5–6 %. Standard deviations of the results of the software measurements were smaller compared to those of the visual estimations. As a result, the proposed procedure would give reliable results with less tested bottles, reducing the material needed as well as test and processing time.

4 Procedure Validation

The proposed procedure was also validated in order to obtain its degree of accuracy. Pixels colors of samples images were manually checked to understand whether the imaging software classified them correctly (i.e. background, bitumen or aggregate) or not.

For this purpose a specific Java-based code was written, in order to superimpose a 51 × 51 grid of equidistant points on both the non-classified and the classified image and extract their RGB values and x, y coordinates (Fig. 4).

The selected pixels of both images were grouped in three different classes (aggregate, bitumen and background) based on their RGB values. According to the described procedure of recognition, the passage from the non-classified to the classified image was analyzed: if the estimation of the software were correct the bitumen classified pixels would be still classified as bitumen, while the aggregate pixels would be re-classified as background. The superimposed grid helped in identifying whether the software classification was made correctly or not.

Figure 5 shows a detail of the pixels' color inspection. Circles represent the correctly classified pixels, (1) is background, (2) is bitumen and (3) is aggregate. The square (4) is an example of misclassification; in fact this pixel was classified as

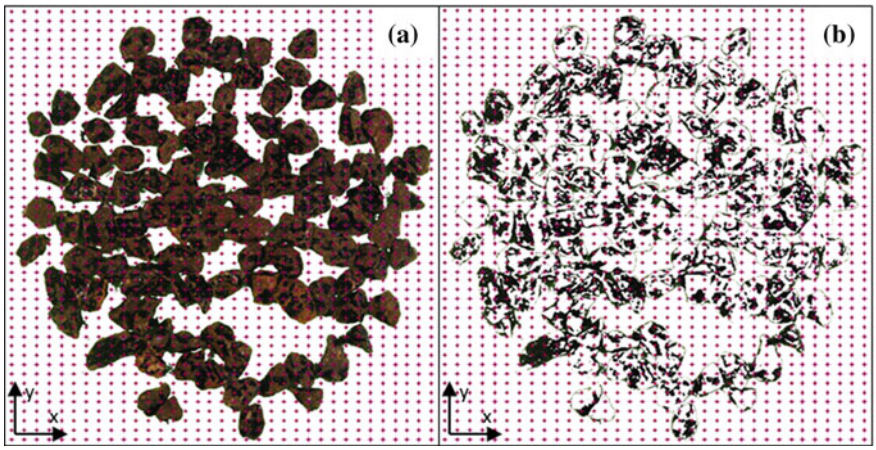

Fig. 4 Validation of the technique: pixels' color inspection for the non-classified (**a**) and classified (**b**) images

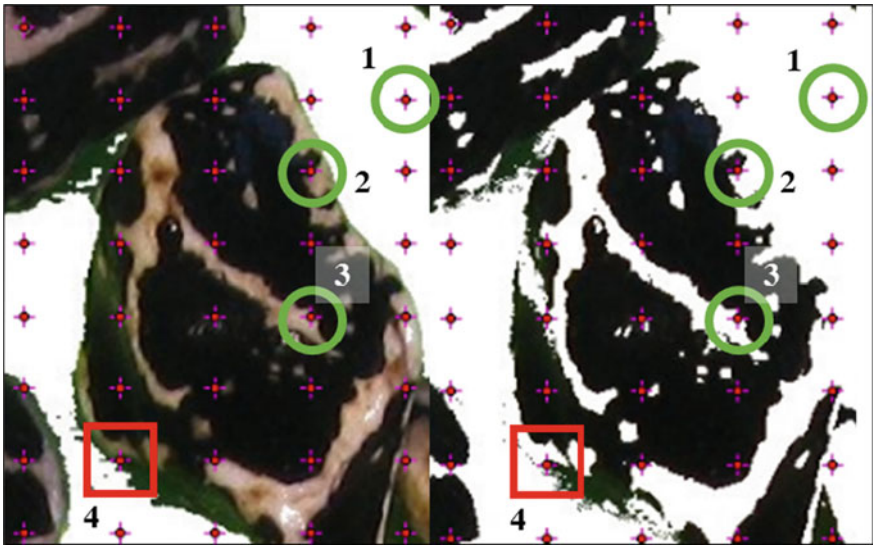

Fig. 5 Detail of pixels' color inspection: correct and incorrect classifications

bitumen while it was actually a shadow pertaining to the background. However, almost all shadows were classified correctly and represented far less than 1 % of all the background pixels.

The adopted procedure validation allowed the generation of confusion matrices that could be used in a series of descriptive and analytical techniques, such as those based on accuracy indices.

In the search for the fundamental characteristics of a generalized confusion matrix for classifications, it is suggested that the matrix should fulfill two characteristics in order to identify a perfect matching case:

- Diagonalization. The matrix should be diagonal if, and only if, the assessed data match perfectly the reference data;
- Marginal sums. Marginal sums should match the total grades from the reference and assessed data.

Four different accuracy indexes were computed for each matrix (Table 2).

Where *Pkj* is the element in row *k* and column *j* and *Pkk* is the element in row *k* and column *k*. An overestimation of the reference pixel membership by the assessed pixel membership leads to errors of commission type (linked to the User accuracy). These commission errors appear in the off-diagonal cells along the row of the class. Conversely, an underestimation of the reference value by the assessed value leads to errors of omission type (linked to the Producer accuracy). These omission errors appear in the off-diagonal cells along the column of the class (Silván-Cárdenas et al. 2008). Tables 3, 4, 5, and 6 show the confusion matrices of the tested aggregates and their accuracy indexes. "*bkGD*" stands for background,

Table 2 Accuracy indexes for a matrix with k classes

Overall accuracy (OA)	Bitumen-aggregate accuracy (BAA)	kth user accuracy (UA) (k)	kth producer accuracy (PA) (k)
$\frac{\sum_k Pkk}{\sum_{kj} Pkj}$	$\frac{\sum_{k=1}^{2} P_{kk}}{\sum_{k,j=1}^{2} P_{kj}}$	$\frac{P_{kk}}{\sum_j P_{kj}}$	$\frac{P_{kk}}{\sum_j P_{jk}}$

Table 3 Confusion matrix and accuracy indexes for Limestone

	Class	Reference data			Sum	User's accuracy (%)
		L	Bit.	bkGD		
Classified image	L	204	17	2	223	91.5
	Bit.	5	529	8	542	97.6
	bkGD	0	0	1836	1836	100.0
Sum		209	546	1846	2601	
Producer's accuracy (%)		97.6	96.9	99.5		

Table 4 Confusion matrix and accuracy indexes for Porphyry

	Class	Reference data			Sum	User's accuracy (%)
		P	Bit.	bkGD		
Classified image	P	421	10	0	431	97.7
	Bit.	15	316	17	348	90.8
	bkGD	0	0	1822	1822	100.0
Sum		436	326	1839	2601	
Producer's accuracy (%)		96.6	96.9	99.1		

Table 5 Confusion matrix and accuracy indexes for Basalt

	Class	Reference data			Sum	User's accuracy (%)
		B	Bit.	bkGD		
Classified image	B	133	16	1	150	88.7
	Bit.	20	568	11	599	94.8
	bkGD	0	0	1852	1852	100.0
Sum		153	584	1864	2601	
Producer's accuracy (%)		86.9	97.3	99.4		

while "*Bit.*" refers to the bitumen. Results show that the software easily drops out the background, with a producer's accuracy of 100 %. The example of Table 3 shows that 5 out of the 208 total pixels of Limestone were classified as bitumen while 17 out of the 222 pixels classified by the software as Limestone actually belongs to bitumen.

Table 6 Confusion matrix and accuracy indexes for Blast furnace slag

	Class	Reference data			Sum	User's accuracy (%)
		S	Bit.	bkGD		
Classified image	S	127	37	0	164	77.4
	Bit.	13	397	10	420	94.5
	bkGD	0	0	2017	2017	100.0
Sum		140	434	2027	2601	
Producer's accuracy (%)		90.7	91.5	99.5		

Table 7 OA and BAA indexes for the tested aggregates

Aggregate	L	P	B	S
OA (%)	98.8	98.4	98.1	97.7
BAA (%)	97.1	96.7	95.1	91.3

The software well classified pixels belonging to L and P having a Producer's accuracy of respectively 97.6 and 96.6 %. As expected B and S, due to their dark grey color, were by far the most confound with the bitumen. Producer's accuracies were respectively 86.9 and 90.7 %. In the example in Table 6, 37 out of the total 164 pixels classified as blast furnace slag by the software, actually belonged to bitumen. Table 7 shows a summary of the Overall Accuracy (OA) and the Bitumen—Aggregate Accuracy (BAA) indexes for all the adopted aggregates.

The OA index shows that the software is highly accurate. The BAA index indicates that the accuracy in recognizing the difference between the bitumen and the aggregate is greater than 95 % for L, P and B and greater than 90 % for S.

It is concluded that the lighter the color of the aggregates, the higher the BAA.

These results confirmed the validity of the procedure compared to the visual estimation, with a consistent reduction of the errors and of the time for the test execution.

5 Conclusions

Based on the experiments and on the image analyses, the following can be summarized and concluded:

- the software estimation is, by far, more accurate compared to visual estimation;
- the standard deviation of the software results between each observation is smaller, demonstrating a low dispersion of the data;
- more discrepancy between visual and software recognition were observed with lighter aggregates. This variability can affect the comparison between different aggregates;

- all the tested aggregates were effectively recognized by the software, although light and middle dark rock types gave more accurate results due to the contrast between the aggregate and the bitumen;
- being the equipment quite simple and the software open source, this procedure is very cost-effective and may substitute or support the standard visual registration;
- as standard deviations of the results obtained by the software are smaller compared to those of the visual estimations, the proposed procedure requires less bottles to be tested, thus reducing the tested material as well as test and processing time.

References

Bhasin A (2006) Development of methods to quantify bitumen-aggregate adhesion and loss of adhesion due to water. Diss. Texas A&M University. http://repository.tamu.edu/bitstream/handle/1969.1/5934/etd-tamu-2006A-CVEN-Bhasin.pdf?sequence=1. Accessed 21 Nov 2014

Dondi G, Simone A, Vignali V, Manganelli G (2012) Numerical and experimental study of granular mixes for asphalts. Powder Technology 232: 31–40. doi: 10.1016/j.powtec.2012.07.057

EN 12697-11 Bituminous mixtures – Test methods for hot mix asphalt – Part 11: Determination of the affinity between aggregate and bitumen

Ford A, Roberts A (1998) Colour space conversions. Westminster University, London, 1-31

Grönniger J, Wistuba MP, Renken P (2010) Adhesion in bitumen-aggregate-systems: New technique for automated interpretation of rolling bottle tests. Road Materials and Pavement Design 11(4): 881-898. doi: 10.3166/RMPD.11.881-898

Grenfell J, Ahmad N, Liu Y, Apeagyei A, Large D, Airey G (2014) Assessing asphalt mixture moisture susceptibility through intrinsic adhesion, bitumen stripping and mechanical damage. Road Materials and Pavement Design 15(1): 131-152. doi: 10.1080/14680629.2013.863162

Hefer AW (2004) Adhesion in bitumen-aggregate systems and quantification of the effects of water on the adhesive bond. Diss. Texas A&M University. http://repository.tamu.edu/bitstream/handle/1969.1/1457/etd-tamu-2004C-CVEN-Hefer.pdf?sequence=1. Accessed 21 Nov 2014

Kallen H, Heyden A, Astrom K, Lindh P (2012) Measurement of bitumen coverage of stones for road building, based on digital image analysis. Paper presented at IEEE Workshop on the Applications of Computer Vision, WACV 2012, Breckenridge, 9 January 2012

Ibraheem NA, Hasan MM, Khan RZ, Mishra PK (2012) Understanding Color Models: A Review. ARPN Journal of Science and Technology 2(3): 265-275. ISSN 2225-7217

Miller JS, Bellinger WY (2003) Distress Identification Manual for the Long-Term Pavement Performance Program. Publication FHWA-RD-03-031. FHWA, Virginia

Mulsow C (2012) Determination of the degree of gravel aggregate-bitumen coverage by multi-directional reflectance measurements. Paper presented at the XXII ISPRS Congress, Melbourne, 25 August – 01 September 2012

Silván-Cárdenas JL, Wang L (2008) Sub-pixel confusion–uncertainty matrix for assessing soft classifications. Remote Sensing of Environment 112(3):1081-1095. doi: 10.1016/j.rse.2007.07.017

Vignali V, Mazzotta F, Sangiorgi C, Simone A, Lantieri C, Dondi G (2014) Rheological and 3D DEM characterization of potential rutting of cold bituminous mastics. Construction and Building Materials 73: 339-349. doi: 10.1016/j.conbuildmat.2014.09.051

http://www.springer.com/978-94-017-7341-6

8th RILEM International Symposium on Testing and Characterization of Sustainable and Innovative Bituminous Materials

Canestrari, F.; Partl, M.N. (Eds.)
2016, XXVI, 1042 p. 551 illus., 352 illus. in color.,
Hardcover
ISBN: 978-94-017-7341-6