MODELLING THE STONE SIZE DISTRIBUTION OF A DIAMOND DEPOSIT

CHRISTIAN LANTUÉJOUL and MIKE MILLAD
Ecole des Mines de Paris, 35 rue Saint Honoré, 77305 Fontainebleau, France.
De Beers MRM, PO Box 23329, Claremont 7735, Cape Town, South Africa.

ABSTRACT

The diamond size-frequency distribution (SFD) is one of the most important factors determining the revenue of a diamond deposit. Since the diamond SFD typically varies spatially in an alluvial deposit, it should be modelled locally. A data set from an alluvial deposit was used to compute experimental simple- and cross-variograms for variables representing diamond abundance in a range of standard granulometric classes. These variograms display a number of striking features, firstly the larger the granulometric class the smaller the range, and secondly a negative correlation between the nugget effect (as a proportion of the total sill) and the proportion of available stones. For the sake of prediction and simulation, a simple model is proposed to reproduce these features, while also taking into account the particulate nature of diamond. At first, all points from a Cox process are interpreted as stones and assigned independent sizes. Then each stone is independently migrated as a function of its size, as would occur by geological processes. The resulting process is a marked Cox process, the properties of which are investigated.

INTRODUCTION

Prediction of the diamond size-frequency distribution (SFD) in a diamond deposit, contributes significantly towards the estimation of expected revenue. Following a case study (Millad, 2007), which involved the estimation of the diamond SFD in an alluvial diamond deposit, it became apparent that standard multivariate models were not the most appropriate. The desired result of the case study was an estimate of diamond SFD, expressed in terms of the number of stones in a range of granulometric classes, called sieve classes, at a panel scale. The particulate nature of diamonds and some properties of the experimental simple- and cross-variograms

prompted the design of a new, more suitable model. The objective of this paper is therefore twofold:

i) To examine the experimental characteristics (statistics and spatial structures) of the data, as used in the case study.

ii) To propose a marked point process for the modelling of the spatial distribution of the stones, including their size attribute, in full compatibility with the experimental observations mentioned in i) above.

EMPIRICAL OBSERVATIONS

Background

The case study area is located several kilometres offshore of the Namibian southwest coast, to the immediate north of the Orange River mouth (Figure 1). This area, called Atlantic 1, contains a major alluvial diamond deposit, which is currently sampled and exploited using vessel-based mining and mineral processing systems. The diamonds are hosted within thin (typically $< 0.5m$) grit and gravel horizons that occur on the seabed, which are treated as two-dimensional entities during the estimation of the diamond resource.
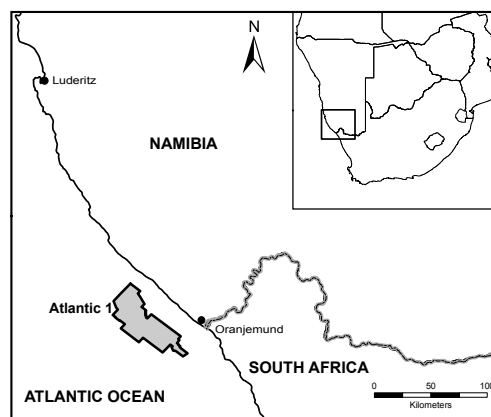

Figure 1: Location map.

The sampling vessel uses an airlift drill system to recover sediment and diamonds from the seabed (sample size $= 12.2m^2$). Once recovered from the sediment, the weight of each diamond or "stone" is measured in units of carats ($1ct = 0.2g$). The stones are also sieved into granulometric classes, called "sieve classes", using a standard set of sieves (see Figure 2). The sieve class into which each stone falls is recorded.

A test area with approximatively $4km \times 4km$ of uninterrupted sample coverage, situated in the southeast of Atlantic 1, was selected for the multivariate analysis. In order to simplify the testing of the multivariate approach, samples were declustered,
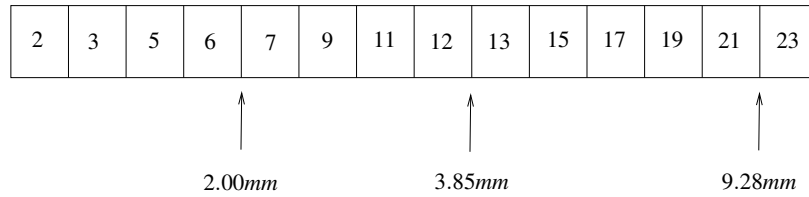
2	3	5	6	7	9	11	12	13	15	17	19	21	23

$2.00mm$ $3.85mm$ $9.28mm$

Figure 2: The family of sieve classes and their key aperture sizes

leaving a regular, $71m \times 71m$ sample grid.

Choice of variables

The sample diamond SFD in Atlantic 1 was expressed in terms of stone density (ie. $stns/m^2$) per sieve class. Such variables are additive, ensuring that the diamond abundance is taken into account when estimating the SFD. This resulted in 9 variables for structural analysis, ranging from the sieve class 7 to 21 (sieve class 23 was excluded due to a lack of observations). Some 3244 samples were available for analysis.

Basic statistics and correlation

Basic statistics and a correlation matrix for the 9 variables are displayed in Tables 1 and 2, respectively. Note that because the number of stones is integer, and sample size is constant at $12.2m^2$, the variables are discrete. The coefficient of variation and skewness are inversely related to the proportion of total stones in each sieve class, while being positively correlated with the proportion of barren samples. These relationships are predictable, and are interpreted as a function of both the inherent variability of stone densities in different size categories (larger stones = greater statistical variability) and the relative abundance of the stones (lower stone densities = greater standardised statistical variability).

Table 1: Basic statistics for the sieve class stone density

class	min	max	mean	std dev	CoV	% zeroes	% stones
7	0	0.984	0.021	0.050	2.45	80.73	8.50
9	0	0.820	0.047	0.083	1.77	64.58	19.45
11	0	1.148	0.052	0.090	1.73	62.39	21.61
12	0	0.656	0.034	0.069	2.02	72.23	14.04
13	0	0.902	0.049	0.093	1.89	64.83	20.31
15	0	0.328	0.013	0.036	2.82	87.02	5.22
17	0	0.410	0.012	0.036	2.88	87.39	5.14
19	0	0.574	0.011	0.038	3.33	89.12	4.69
21	0	0.164	0.002	0.013	6.48	97.60	0.85

Table 2: Correlation matrix for the sieve classes

	7	9	11	12	13	15	17	19	21
7	1.00	0.30	0.26	0.21	0.24	0.09	0.15	0.10	0.04
9	0.30	1.00	0.46	0.38	0.38	0.22	0.27	0.28	0.11
11	0.26	0.46	1.00	0.44	0.50	0.29	0.29	0.30	0.13
12	0.21	0.38	0.44	1.00	0.46	0.26	0.30	0.30	0.11
13	0.24	0.38	0.50	0.46	1.00	0.28	0.35	0.31	0.14
15	0.09	0.22	0.29	0.26	0.28	1.00	0.26	0.22	0.10
17	0.15	0.27	0.29	0.30	0.35	0.26	1.00	0.25	0.08
19	0.10	0.28	0.30	0.30	0.31	0.22	0.25	1.00	0.13
21	0.04	0.11	0.13	0.11	0.14	0.10	0.08	0.13	1.00

The correlation coefficients between the 9 variables are all positive. These positive relationships were expected, since an increase in overall stone density will generally be echoed within each sieve class. Most of the variables also show weak-to-moderate correlations with their counterparts, with the best correlation of 0.5 between the sieve classes 11 and 13. It is noteworthy that there is good correspondence between the general degree of correlation that a variable enjoys with its 8 partners and the proportion of the total available stones supporting that variable.

Experimental simple- and cross-Variograms

The size of the multivariate system prohibits illustration of all the simple- and cross-variograms. Therefore, Figure 3 shows part of the system, for the sieve classes 9 through 13. This is sufficient to illustrate the general features of the spatial interdependence between sieve classes. Note that no evidence of anisotropy could be found for any of the variables, hence the analysis using omnidirectional variograms.

The following features characterise the simple- and cross-variograms:

i) Very high nugget effects in the simple-variograms, this being reflective of the small sample support size for the individual sieve classes. This is not surprising, given that the sample size was originally optimised for the estimation of total stone density across all sieve classes. The nugget effect is also inversely related to the sample mean and the proportion of stones supporting each variable, and therefore is largely a function of the amount of information available for each sieve class.

ii) The ranges of simple-variograms generally decrease with increasing stone size, and are therefore distinct from one another. This is believed to be due in part to the reduced mobility of stones in the hydraulic environment, coupled with a general decrease in stone abundance, as stone size increases.

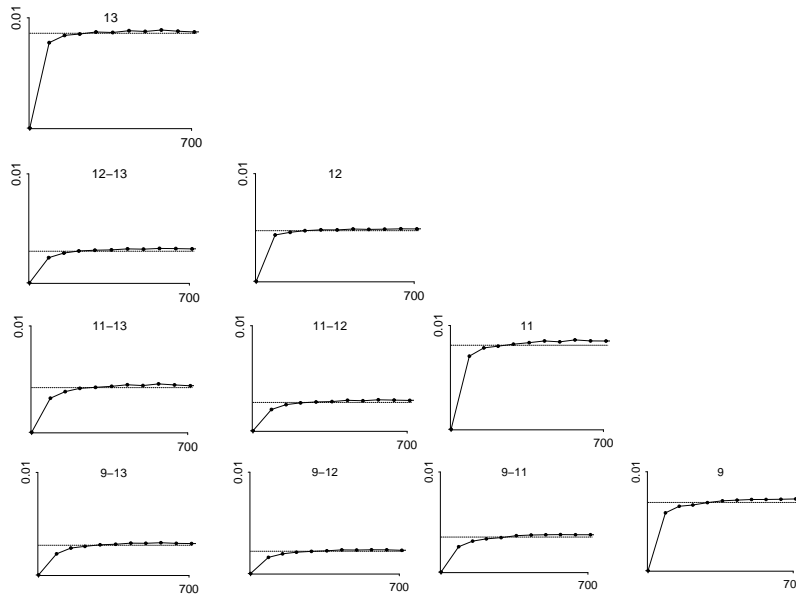iii) Although the experimental cross-variograms have a similar shape to the

Figure 3: Simple- and cross-variograms for the sieve classes 9 through 13.

simple-variograms, they are nonetheless more difficult to interpret. Debatably, the longer range component of the cross-variograms is somewhat more important than in the simple-variograms. This is in accordance with the fact that particles are size-sorted by hydraulic processes and lends support to this inference.

THE MODEL

Fitting simple- and cross-variograms becomes progressively more difficult as the number of granulometric classes increases. This is especially true for large classes that contain a limited number of stones but which represent a substantial portion of the value of the deposit. It has, however, long been observed that there exist important dependence relationships between data from neighbouring granulometric classes. Conversely, it would be quite helpful if these relationships could contribute to the structural modelling. As a consequence, rather than fitting the structure using standard models, why not consider a stochastic model, based on some simple and plausible physical process, and which is capable of reproducing the observed dependence relationships?

The model proposed below is a first attempt toward this approach. After a short reminder about the Cox process, a migration model is introduced and its properties investigated. To make the text more flowing, all proofs have been relegated to the Appendix.

Reminder on the Cox process

A Cox process is a Poisson point process with a random intensity function. This random function is denoted by Z and called *potential*. It indicates the propensity of the points to be preferentially located in some regions rather than others. The top left of Figure 4 shows a realisation of a Cox process superimposed on its potential.

In order to characterise the statistical properties of a point process, we can consider the functional that assigns to each domain A the probability $P\{N(A) = 0\}$ that it is devoid of points. Originated from the random set theory (Matheron, 1975), this functional acts for random sets exactly as a complementary distribution function for random variables. This is the reason why we call it the *complementary distribution function* of the point process, or *cmdf* for short. The cmdf of a Cox process with potential Z is

$$P\{N(A) = 0\} = E\left\{e^{-Z(A)}\right\} \tag{1}$$

where $Z(A)$ denotes the integral of Z over the domain A (Kleingeld and Lantuéjoul, 1993).

Suppose that Z is second order stationary with mean m and covariance C_Z, and let A be a bounded domain of $I\!R^d$. Then the random function $N(A_.) = \left(N(A_x), x \in I\!R^d\right)$ is also second order stationary with mean $m|A|$ and covariance[1]

$$C_{N(A_.)}(h) = C_Z \star K_A(h) + m K_A(h) \tag{2}$$

where $K_A(h) = |A \cap A_h|$ is the geometric covariogram of A. This covariance is therefore a sum of two terms. The first term is a long range structure conveyed by the potential. The second term is a short range structure that stems from the Poisson seeding of the points once the potential has been fixed. This short structure induces an apparent nugget effect, whose amplitude can be expressed as a proportion of the total variance

$$\theta_{N(A_.)} = \frac{m K_A(0)}{C_{N(A_.)}(0)} = \frac{m|A|}{C_Z \star K_A(0) + m|A|} \tag{3}$$

A migration model

The starting point is a Cox process. Each point of the process, or particle, is assigned a random size, then randomly migrated w.r.t. its size (cf Figure 4).

By calculating its cmdf, it can be shown (cf. Appendix) that this migration process is also a Cox process, with potential

$$\bar{Z}(x) = Z \star g(x) \tag{4}$$

---

[1]In this formula, the symbol $\star$ between C_Z and K_A is a convolution product. Explicitly

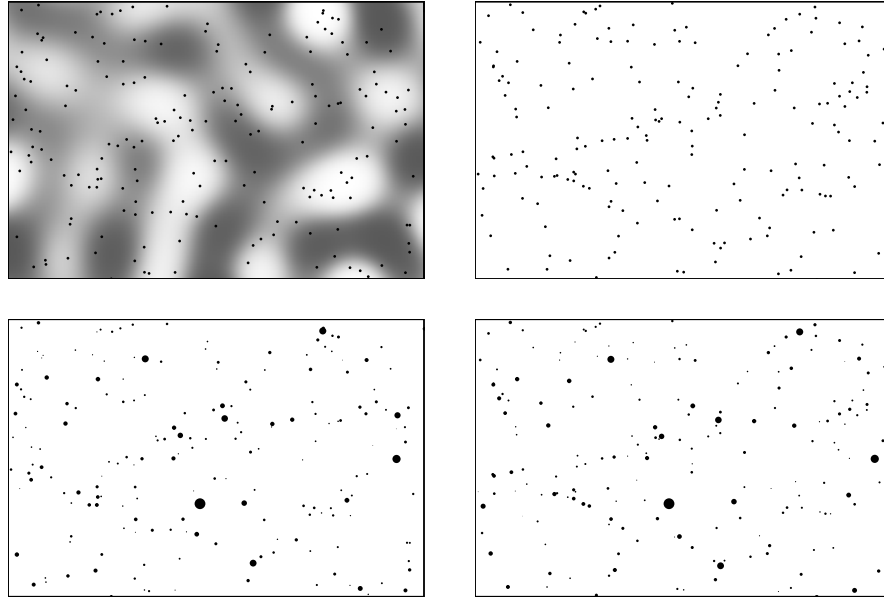$$C_Z \star K_A(h) = \int_{I\!R^d} C_Z(x) K_A(h - x) \, dx$$

.

Figure 4: Model construction. Top left, a realisation of the initial Cox process, superimposed on its potential. Bottom left, the size assignment. Bottom right, the particle migration. Top right, a realisation of the model. In this example, a particle of size s and located at (x,y) is tranported from west to east to (x',y'), where x' and y' are independently normally distributed with respective means $x+m$ and y (m is exponentially distributed with mean $20/(s+0.1)$), and respective variances $10/(s+0.1)$ and $5/(s+0.1)$. As a comparison, the size of the simulation field is 300×200

where g is the migration pdf of a typical particle, obtained by weighting the migration pdf g_s associated with each size s by its relative abundance $f(s)$:

$$g(x) = \int_0^\infty f(s) g_s(x) \, ds \qquad (5)$$

Let $\bar{N}(A_x)$ be the number of points of the migration process in the domain A located at x. According to (2) and (4), the covariance of $\bar{N}(A.)$ takes the form

$$C_{\bar{N}(A.)}(h) = C_{\bar{Z}} \star K_A(h) + mK_A(h) = C_Z \star K_g \star K_A(h) + mK_A(h) \qquad (6)$$

where $K_g = g \star \check{g}$ denotes the transitive covariogram of g (with $\check{g}(x) = g(-x)$). This covariance has the same short structure as the initial Cox Process, but its long range structure has been extended. Note also that $C_Z \star K_g \star K_A(0) \leq C_Z \star K_A(0)$, which implies that

$$\theta_{\bar{N}(A.)} = \frac{m|A|}{C_Z \star K_g \star K_A(0) + m|A|} \geq \frac{m|A|}{C_Z \star K_A(0) + m|A|} = \theta_{N(A.)} \qquad (7)$$

In other words, migration makes the apparent nugget effect increase.

The process associated with one particular granulometric class

In this section we are interested in the particles pertaining to a given granulometric class $S \subset]0, \infty[$. We call them S-particles for short.

Because the model lets the particles migrate independently, the S-particles can be studied irrespective of the other particles. Accordingly, the general results established in the previous section are directly applicable to the S class, as long as the potential Z has been replaced by the potential Z_S of the S-particles

$$Z_S(x) = f(S)Z(x) \tag{8}$$

and the migration p.d.f. g has been replaced by the migration p.d.f. of a typical S-particle

$$g_S(x) = \int_S \frac{f(s)}{f(S)} g_s(x) \, ds \tag{9}$$

Consequently, the migrated S-particles are spatially distributed like a Cox process with potential

$$\bar{Z}_S(x) = Z_S \star g_S(x) = f(S)Z \star g_S(x) \tag{10}$$

and because of (2) the regularised process $\bar{N}_S(A.)$ takes on the covariance

$$C_{\bar{N}_S(A.)}(h) = f^2(S)C_Z \star K_{g_S} \star K_A(h) + f(S)mK_A(h) \tag{11}$$

Note that this covariance depends on the migration p.d.f. g_S via its transitive covariogram G_S. Accordingly, $C_{\bar{N}_S(A.)}$ is affected by the dispersion of g_S, not by its mean. The more scattered g_S, the longer the range of $C_{\bar{N}_S(A.)}$. This is in agreement with empirical observations, insofar as small particles have greater mobility and tend to disperse more than large ones.

Formula (11) also suggests that the apparent nugget effect of $\bar{N}_S(A.)$

$$\theta_{\bar{N}_S(A.)} = \frac{m|A|}{f(S)C_Z \star K_{g_S} \star K_A(0) + m|A|} \tag{12}$$

is a monotonic decreasing function of the class proportion. However, this has to be counterbalanced by the fact that the variance of the long range structure also depends on $f(S)$. Note however that the inequality $C_Z \star K_{g_S} \star K_A(0) \leq C_Z \star K_A(0)$ holds, which implies

$$\theta_{\bar{N}_S(A.)} \geq \frac{m|A|}{f(S)C_Z \star K_A(0) + m|A|}$$

Accordingly, the smaller the class proportion, the larger the minimal value allowed for $\bar{N}_S(A.)$. Thus, what the model implies for the apparent nugget is compatible with the empirical observations without fully corroborating them.

The biprocess associated with two granulometric classes

Let T be another granulometric class such that $T > S$ (short notation for $t > s$ for each $s \in S$ and $t \in T$). What are the dependence relationships between $\bar{N}_S(A.)$ and $\bar{N}_T(A.)$?

At first, it should be mentioned that both processes are dependent, even if they rest on different granulometric classes. This is because both are directed by the same potential Z. Their joint cmdf cannot be factorised:

$$P\{\bar{N}_S(A) = 0, \bar{N}_T(B) = 0\} = E\left\{exp\left[-f(S)Z \star g_S(A) - f(T)Z \star g_T(B)\right]\right\} \quad (13)$$

The cross-covariance of the regularised biprocess takes the form

$$C_{\bar{N}_S(A.),\bar{N}_T(A.)}(h) = f(S)f(T)C_Z \star K_{g_S,g_T} \star K_A(h), \quad (14)$$

where $K_{g_S,g_T} = g_S \star \check{g}_T$ is the transitive cross-covariogram of g_S and g_T. This calls for several comments:

– in contrast to simple-covariances, there is no short range contribution;

– the presence of the convolution product between g_S and $\check{g}_T$ indicates that the cross-covariance is affected by the difference between the mean migrations of classes S and T. This difference acts as a shift factor. In particular, one must expect

$$C_{\bar{N}_S(A.),\bar{N}_T(A.)}(-h) \neq C_{\bar{N}_S(A.),\bar{N}_T(A.)}(h)$$

Of course, this should not be confused with $C_{\bar{N}_S(A.),\bar{N}_T(A.)}(-h) = C_{\bar{N}_T(A.),\bar{N}_S(A.)}(h)$, which always holds;

– in contrast to this, the cross-variogram is symmetric. It does not depend on the difference of migration means, which can be seen directly from its very definition;

– if $S < T$, then g_T is less scattered than g_S. The presence of the convolution product $g_S \star \check{g}_T$ shows that the range of $C_{\bar{N}_S(A.),\bar{N}_T(A.)}$ lies between those of $C_{\bar{N}_S(A.)}$ and $C_{\bar{N}_T(A.)}$.

CONCLUSIONS

A marked Cox point process model was developed following observations on a multivariate data set for an alluvial diamond deposit. This marked Cox process is considered to be more appropriate for modelling the experimental characteristics of such a data set, due primarily to its ability to represent the particulate nature of diamonds. The new model could be used for interpolation or simulation of stone densities in a number of size categories.

REFERENCES

Kleingeld, WJ and Lantuéjoul, C (1993). *Sampling of orebodies with a highly dispersed mineralization*. In A Soares, ed., *Geostatistics Tróia'92*, vol. 2. Kluwer Academic Publisher, Dordrecht, pp. 953–964.

Matheron, G (1975). *Random sets and integral geometry*. Wiley, New York.

Millad, MG (2007). *Testing of an alternative method for estimating diamond size distribution (Atlantic 1, Namibia)*. Tech. rep., Ecole des Mines, Fontainebleau.

APPENDIX

Statistical characterisation of the migration process

Let D be a bounded domain of $\mathbb{R}^d$. By conditioning first w.r.t. the potential and then w.r.t. the content of D, the probability that no particle originating in D has migrated to the domain A can be expressed as

$$
\begin{aligned}
P\{\bar{N}_D(A) = 0\} &= E\left\{ e^{-Z(D)} \frac{Z^n(D)}{n!} \left[\int_D \frac{Z(x)}{Z(D)} \int_0^\infty f(s)[1 - g_s(A_{-x})]dsdx \right]^n \right\} \\
&= E\left\{ \exp\left(-\int_D Z(x) \int_0^\infty f(s)g_s(A_{-x})dsdx \right) \right\}
\end{aligned}
$$

The complementary distribution function of the migrated process is then obtained by letting D tend to $\mathbb{R}^d$:

$$
P\{\bar{N}(A) = 0\} = E\left\{ \exp\left(-\int_{\mathbb{R}^d} Z(x) \int_0^\infty f(s)g_s(A_{-x})dsdx \right) \right\}
$$

Note that it can be simplified by introducing the migration p.d.f. (5)

$$
\begin{aligned}
P\{\bar{N}(A) = 0\} &= E\left\{ \exp\left(-\int_{\mathbb{R}^d} Z(x)g(A_{-x})dsdx \right) \right\} \\
&= E\left\{ \exp\left(-\int_{\mathbb{R}^d} Z(x) \int_A g(y - x)dydx \right) \right\} \\
&= E\left\{ \exp\left(-\int_A Z \star g(y) \right) \right\}
\end{aligned}
$$

Cross-covariance of the regularised biprocess

Here, the purpose is to calculate $C_{\bar{N}_S(A.),\bar{N}_T(A.)}(h)$. By conditioning w.r.t. Z, this covariance can be written as $C_1 + C_2$, with

$$
\begin{aligned}
C_1 &= Cov\{E\{\bar{N}_S(A) \mid Z\}, E\{\bar{N}_T(A_h) \mid Z\}\} \\
C_2 &= E\{Cov\{\bar{N}_S(A), \bar{N}_T(A_h) \mid Z\}\}
\end{aligned}
$$

As $E\{\bar{N}_S(A) \mid Z\} = \bar{Z}_S(A) = f(S)Z \star g_S(A)$ by formula (10), and similarly $E\{\bar{N}_T(A_h) \mid Z\} = f(T)Z \star g_T(A_h)$, we have

$$
C_1 = Cov\{f(S)Z \star g_S(A), f(T)Z \star g_T(A_h)\} = f(S)f(T)C_Z \star K_A \star g_S \star \check{g}_T(h)
$$

On the other hand, $C_2 = 0$ because $\bar{N}_S(A)$ and $\bar{N}_T(A_h)$ are conditionally independent given Z. The covariance is therefore equal to C_1, namely

$$
C_{\bar{N}_S(A.),\bar{N}_T(A.)}(h) = f(S)f(T)C_Z \star K_A \star g_S \star \check{g}_T(h)
$$