

	<b>MAPP</b>	
--	-------------	--

**Titel:**           **Generic Algorithms for Level 3 Products**

**Project:**       MAPP

**Doc. No.:**     MAPP-ATBD-GENL3

**Issue:**        1

**Revision:**    1

**Date:**        8.2.2000

	<u>Function</u>	<u>Name</u>	<u>Organisation</u>	<u>Signature</u>	<u>Date</u>
<b>Author:</b>		Norman Fomferra	SCICON		
		Carsten Brockmann	SCICON		

	<b>MAPP</b>	<b>Doc : MAPP-ATBD-GENL3</b> <b>Projekt : MAPP</b> <b>Name : Generic Level 3</b> <b>Ausg. : 1            Rev : 1</b> <b>Datum : 8.2.2000 Seite: ii</b>
--	-------------	--

### Internal Distribution

<u>Name</u>	<u>Organisation</u>	<u>Quantity</u>
-------------	---------------------	-----------------

### External Distribution

<u>Name</u>	<u>Organisation</u>	<u>Quantity</u>
-------------	---------------------	-----------------

### Change Record

<u>Issue</u>	<u>Revision</u>	<u>Date</u>	<u>Changes</u>
draft		25.08.1998	initial issue
1	0	15.02.1999	Integration concepts
1	1	08.02.2000	Appendix 4.2 and 4.3 added

## 1 Abstract

This document discusses some ways how the L3 data product generating of the Meris VA Processor can be executed in an efficient way.

## 2 Introduction

In generating L3 data products with the Meris VA Processor, pixel containing valid L2 data will be mapped to a fixed spatial grid whose resolution elements have an equal area. These square *bins* are arranged in a square matrix, such that an area of approximately  $3840 \times 3840$  km is covered. The geographical extension of the grid is shown in Figure 1: .

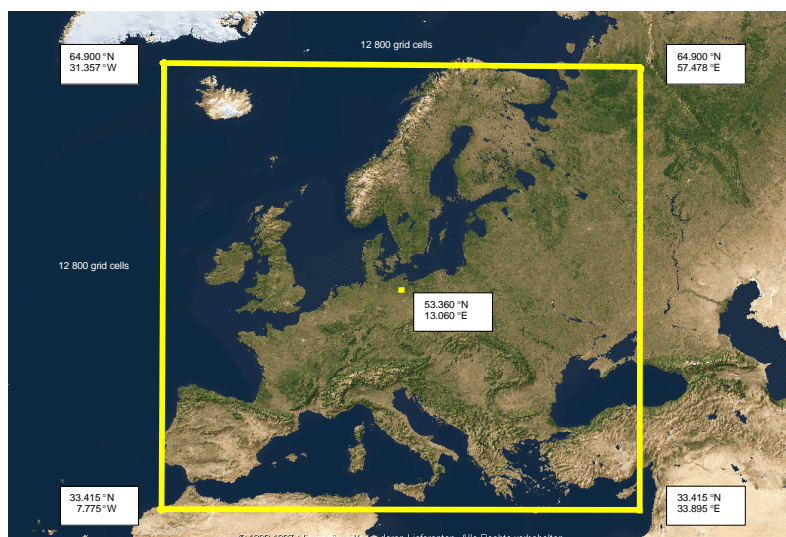


Figure 1: MERIS-VA Level 3 grid

The following spatial grids will be provided in the Meris VA Processor:

Binning Matrix	Resolution
M <sub>1</sub>	12800 × 12800
M <sub>2</sub>	6400 × 6400
M <sub>4</sub>	3200 × 3200
M <sub>8</sub>	800 × 800
M <sub>16</sub>	400 × 400

For each L3 product, multiple matrices must be stored to be able to derive the statistical measures as mean, standard deviation, median and mode.

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8. February 2000	4-11
--------	--	-------------------------------------	------

### 3 Details

#### 3.1 Binning

##### 3.1.1 Mapping

The Standard ESA MERIS Level 1b and Level 2 products are geo – referenced on a 16x16 pixel sub-grid (the so called Tie-points). By interpolation, the geographical co-ordinates for every pixel can be derived with an accuracy of less than 1 pixel (pers. comm. S. Delwart, ESA).

An azimuthal equal area transformation fulfils the requirement of equal grid sizes of 300m everywhere and of centering the grid at the receiving station Neustrelitz. In the annex, the routine which implements that mapping transformation is specified.

##### 3.1.2 Implementation considerations

A single  $M_1$  matrix will have the size of 625 MB for bin elements of 4 bytes. In order to access multiple of such large files per over-flight processing in an efficient way, an examination has been carried out to compare the performance of the UNIX standard library I/O routines and the file mapping routines of the UNIX system interface (TN SW 2). It has been found out that file mapping is about 20 times faster than the standard file I/O.

Although file mapping is faster, it has the disadvantage that the size of all simultaneous mapped regions together must be found somewhere in the addressable memory range, which was 2 GB at the time the architecture of the Meris VA Processor has been established.

In order to make the best use of the file mapping mechanism, only parts of the binning matrices can be mapped simultaneously. In the following, a few suggestions concerning an efficient L3 data processing are given:

1. A single over-flight processing generates all L3 data products with all binning matrices partly mapped and re-mapped at the same time.

**Disadvantages:** The scan line direction of the over-flight (swath) is not parallel to the direction of the binning matrix scan lines. This may require multiple file re-mappings for each swath being processed, which may drastically reduce the time benefit of file mapping in general. Encouraged is that effect by too small memory blocks, because many file must be mapped at the same.

2. A single over-flight processing generates only one L3 data product at the same time with partly or completely mapped binning matrices.

**Disadvantages:** Same as under 1., but allows larger memory blocks being mapped at the same time.

3. Same as 1. or 2., but transforms every pixel in a way that the swath direction is parallel to the direction of the binning matrix scan lines.

**Advantages:** Compensates the disadvantages of 1. and 2.

**Disadvantages:** Unknown influence and acceptance of the transformation concerning the statistics yielded by the binning algorithms.

4. A single over-flight processing generates only parts of all L3 data products with a fixed part of the binning matrices being mapped.

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8. February 2000      5-11
--------	---	---

**Advantages:** There is no file re-mapping required, because the mapped file region stays the same for all binning matrices.

**Disadvantage:** The fixed memory blocks may be too small, because many matrices have to be mapped at the same time, which would in turn require multiple over-flight processings to complete the generation of all L3 data products.

5. A single over-flight processing generates only the part of a single L3 data product with a fixed part of the binning matrices being mapped.

**Advantages:** Same as under 3., but in addition the larger memory blocks being mapped for a single L3 data product cover more matrix scan lines at the same time.

**Disadvantages:** Multiple over-flight processings necessary to complete the generation of a single L3 data product.

### 3.2 Integration

Level 3 data are temporal and spatial integrated Level 2 products. This integration is a statistical process, which needs to be defined according to the answers to the following questions:

- In which way will the Level 3 data be used. Is it foreseen to generate Level 4 data (i.e. integration into models)?
- Which distributions have the Level 2 parameters? Which parameters characterise these distributions?
- Which statistical parameters need to be stored together with the Level 3 product? On which grid?
- In a temporal series, shall all data be equally weighted even if they are unequally distributed?

#### 3.2.1 Spatial Integration

These questions make clear that it is important to recognise the distribution of the Level 2 parameter. Most common are the normal distribution (applies generally to random process) and the log-normal distribution (applies often to biological processes).

The mean value of a normal distribution

$$\bar{X}_{avg} = \frac{1}{n} \sum_{i=1}^n x_i$$

leads to an underestimation of the real mean if the sample size is small.

In the case of the log-normal distribution, the values are log-transformed, and then the statistical parameters are calculated. The mean is then retrieved by exponential back-transformation. In the case of the mean value, this is the geometrical mean.

$$m_x = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$$

$$\bar{X}_{geom} = e^{m_x}$$

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8. February 2000	6-11
--------	---	-------------------------------------	------

In the case of distributions with a positive slope, the geometrical mean (= median) underestimates the real mean value.

A generally good estimate for normal and log-normal distributions, for a large variance and for also for small sample size gives the Maximum-Likelihood estimation (MLE):

$$\bar{X}_{mle} = e^{\left(m_x + \frac{1}{2}s_x^2\right)}$$

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n [\ln(X_i) - m_x]^2$$

In the course of SeaWiFS algorithm definition, the NASA has found for satellite derived chlorophyll distributions that

- the arithmetic and the MLE mean are in good agreement
- the geometrical mean shows significant deviations from the arithmetic mean
- the MLE is best suited for further Level 4 data generation. It allows to derive the mean of an L4 variable from mean L3 variables, and covariances can easily be retrieved from MLE parameters

### 3.2.2 Temporal integration

The temporal domain is usually very irregularly filled. Because of revisit time and cloud coverage, in a given period of, for example one week, various different numbers of data can be available. In addition, due to the spatial binning, more than one pixel of a single overpass can fall into one bin, so that the daily values already can have different statistics. The most easy way to handle, which in fact was applied to CZCS, is to mosaic the images, i.e., to define a criterion (e.g. highest value) which is applied to every bin. The above mentioned differences in data series are disregarded. This is a very arbitrary method which should not applied here.

The next easy solution is to apply a weighted mean and then the main question is how to define the weights which will be assigned to the individual data values. These weights, w, should depend on the numbers of values, n (i denotes the bin index):

$$w_i = \frac{1}{n_i^p}$$

The exponent p determines the behaviour of the weighting:

- p=0 (no weighting) gives too much weight to “good” days, i.e. those days with many data
- p=1 (arithmetic mean) gives too much weight to “bad” days, i.e. those days with only one measurement
- p=0.5 is the compromise which was chosen for SeaWiFS and which is proposed for MAPP

### 3.3 MAPP Algorithm

The following sums of the data and of the squares of the data are calculated:

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8.February 2000 7-11
--------	--	---

$$S_1 = \sum_{i=1}^N \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \ln(X_{ij})$$

$$S_2 = \sum_{i=1}^N \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} [\ln(X_{ij})]^2$$

$X_{ij}$  is the  $i^{\text{th}}$  observation of the Level 2 variable  $X$  at the time  $t_j$

$n_i$  is the number of observations at time  $t_j$

It follows, that

$$n = \sum_{i=1}^N n_i$$

is the total number of observation in a bin, and

$$W = \sum_{i=1}^N \sqrt{n_i}$$

is the sum of the weights. The Level 3 products (i.e., statistics) are then calculated:

$$m_x = \frac{S_1}{W} \quad \text{and} \quad s_x^2 = \frac{S_2}{W} - m_x^2$$

$$\bar{X}_{mle} = e^{(m_x + \frac{1}{2}s_x^2)}$$

$$SD_X = \bar{X}_{mle} \cdot \sqrt{e^{s_x^2} - 1}$$

$$\bar{X}_{med} = e^{m_x}$$

$$\bar{X}_{mod} = e^{m_x - s_x^2}$$

$m_x$  and  $s_x$  are only defined for the following calculations.  $\bar{X}_{mle}$  is the mean value,  $SD_X$  is the standard deviation,  $\bar{X}_{med}$  is the median and  $\bar{X}_{mod}$  the modus.

With these statistics it is possible to calculate exactly Level 4 variable for some common cases:

Linear model:

$$Y = A + BX$$

$$\bar{Y}_{mle} = A + B\bar{X}_{mle}$$

$$SD_Y = B \cdot SD_X$$

Power model:

$$Y = A + X^B$$

$$m_y = \ln(A) + Bm_x$$

$$s_y^2 = B^2 \cdot s_x^2$$

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8.February 2000	8-11
--------	--	------------------------------------	------

## 4 Appendix:

### 4.1 Mapping Routine

dMax = 1920

dMin = -1920

dR = 6372

m = <Number of Grid Elements, e.g. 12800>

x = dMin + (double)i / (double)(m) \* (dMax - dMin);

y = dMax - (double)j / (double)(m) \* (dMax - dMin);

dLam = x / (dR \* Math.cos(y / dR));

dPhi = y / dR;

Coord coord = pl.getCoordAt(j);

coord.x = dR \* Math.cos(dPhi) \* Math.cos(dLam);

coord.y = dR \* Math.cos(dPhi) \* Math.sin(dLam);

coord.z = dR \* Math.sin(dPhi);

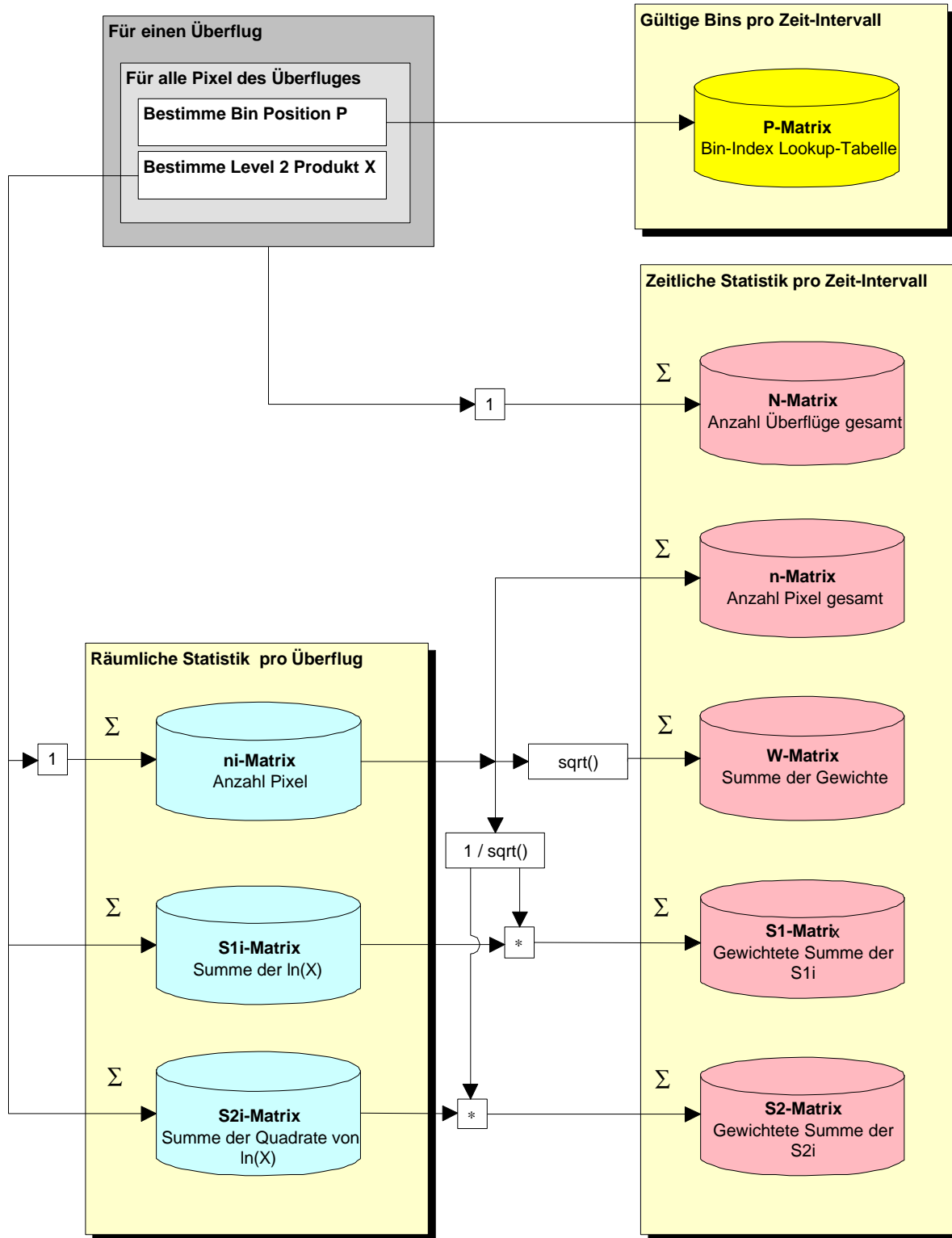
coord.rotateY(-dPhi0);

coord.rotateZ(dLam0);

```
public void rotateY(double ay) {
    double cosa = Math.cos(ay);
    double sina = Math.sin(ay);
    double xx = x * cosa + z * sina;
    double zz = z * cosa - x * sina;
    x = xx;
    z = zz;
}
```

```
public void rotateZ(double az) {
    double cosa = Math.cos(az);
    double sina = Math.sin(az);
    double xx = x * cosa - y * sina;
    double yy = y * cosa + x * sina;
    x = xx;
    y = yy;
}
```

### 4.2 Data Flow Concept



SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8. February 2000	10-11
--------	--	-------------------------------------	-------

### 4.3 Transformation Geodetic Coordinates into geographical coordinates

a, b  $\Rightarrow$   $\lambda$ ,  $\varphi$

1. Transformation from geodetic map co-ordinates a,b to geographical co-ordinates  $\lambda'$ ,  $\varphi'$ :

$$I' = \frac{a}{r \cos \frac{b}{r}}, \quad j' = \frac{b}{r}$$

2. Transformation from to geographical co-ordinates  $\lambda'$ ,  $\varphi'$  to global co-ordinates  $x'$ ,  $y'$ ,  $z'$ :  
(The positive z-axis of the global co-ordinate system is given by the direction earth centre to north pole.)

$$x' = r \cos j' \cos I', \quad y' = r \cos j' \sin I', \quad z' = r \sin j'$$

3. Transformation to the projection centre point  $\lambda_0, \varphi_0$  given by a rotation around the y-axis with with  $\varphi_0$  degrees and a subsequent rotation around the z-axis with  $\lambda_0$  degrees:

$$x = k \cos I_0 - y' \sin I_0, \quad y = y' \cos I_0 + k \sin I_0, \quad z = z' \cos(-j_0) + x' \sin j_0$$

with  $k = x' \cos(-j_0) - z' \sin j_0$

4. Transformation from global co-ordinates x,y,z to geographical co-ordinates  $\lambda'$ ,  $\varphi'$  to the projection centre point  $\lambda_0, \varphi_0$  given by a rotation around the y-axis with with  $\varphi_0$  degrees and a subsequent rotation around the z-axis with  $\lambda_0$  degrees:

$$I = \arcsin \frac{y}{\sqrt{x^2 + y^2}}, \quad j = \arcsin \frac{z}{r}$$

4.3.1 Transformation:  $\lambda$ ,  $\varphi$   $\rightarrow$  a, b

$$x = r \cos j \cos I, \quad y = r \cos j \sin I, \quad z = r \sin j$$

$$x' = k \cos j_0 + z \sin j_0, \quad y' = y \cos(-I_0) - x \sin I_0, \quad z' = z \cos j_0 - k \sin j_0$$

with  $k = x \cos(-I_0) + y \sin I_0$

$$I' = \arcsin \frac{y'}{\sqrt{x'^2 + y'^2}}, \quad j' = \arcsin \frac{z'}{r}$$

SCICON	Generic Algorithms for Level 3 Products	MAPP-ATBD-GENL3 8.February 2000      11-11
--------	--	---

$$a = l \cdot r \cdot \cos j', \quad b = j' \cdot r$$

a [km]	b [km]	l [deg]	j [deg]
0	0	13.060	53.360
-1920	1920	-31.357	64.900
-1920	-1920	-7.775	33.415
1920	1920	57.478	64.900
1920	-1920	33.895	33.415