

*DRAFT DOCUMENT (March 2015)  
Prepared by AC Le Gall (ICP M&M)  
and Max Posch (CCE).*

*To be reviewed by ICP M&M NFCs*

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

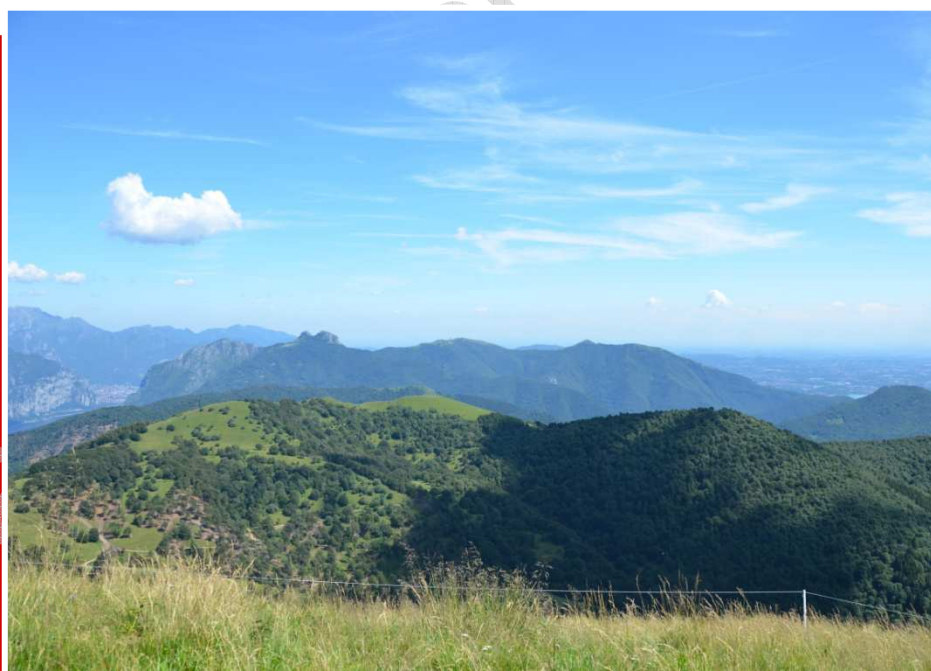
Chapter VI

Chapter VII

**Chapter VIII**

Annexes

## VIII. GENERAL MAPPING ISSUES



Updated by Max Posch and Anne Christine Le Gall, Chairwoman of the Task Force on Modelling and Mapping from initial text by M. Posch and J. Aherne (Mapping Manual 2004 and various CCE reports) and from documentation on EMEP.

Draft (March 2015)

Please refer to this document as: CLRTAP, 2015. General mapping issues, Chapter VIII of Manual on methodologies and criteria for modelling and mapping critical loads and levels and air pollution effects, risks and trends. UNECE Convention on Long-range Transboundary Air Pollution; accessed on [date of consultation] on Web at [www.icpmapping.org](http://www.icpmapping.org).”

Update: 2 March 2015

## TABLE OF CONTENT

<b>VIII. GENERAL MAPPING ISSUES.....</b>	<b>1</b>
<b>VIII.1 GEOGRAPHIC GRID SYSTEMS.....</b>	<b>5</b>
VIII.1.1 <i>The polar stereographic projection.....</i>	5
VIII.1.2 <i>Grids used for critical loads mapping under LRTAP Convention.....</i>	6
VIII.1.2.1 <i>The latitude-longitude grid.....</i>	6
VIII.1.2.2 <i>The 50x50 km<sup>2</sup> grid (EMEP50 grid).....</i>	7
VIII.1.2.3 <i>The 150x150 km<sup>2</sup> grid (EMEP150 grid).....</i>	8
VIII.1.3 <i>The area of an EMEP grid cell.....</i>	10
<b>VIII.2 COMPARING CRITICAL LOADS: CUMULATIVE DISTRIBUTION FUNCTIONS, PERCENTILES AND PROTECTION ISOLINES.....</b>	<b>12</b>
VIII.2.1 <i>Cumulative distribution function.....</i>	12
VIII.2.2 <i>Quantiles and percentiles.....</i>	13
VIII.2.2.1 <i>Linear interpolation of the cdf.....</i>	13
VIII.2.2.2 <i>Empirical distribution function:.....</i>	15
VIII.2.3 <i>Percentile functions and protection isolines.....</i>	16
<b>VIII.3 CRITICAL LOAD EXCEEDANCES USED IN INTEGRATED ASSESSMENT MODELLING.....</b>	<b>18</b>
VIII.3.1 <i>Gap Closure Methods.....</i>	18
VIII.3.2 <i>LINEAR emission-exceedance relationships.....</i>	21
<b>VIII.4 REFERENCES.....</b>	<b>21</b>

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

Annexes



In this Chapter procedures are described for summarising results of critical load and exceedance calculations and presenting them on a regional scale. The objective of the text below is to show the (mathematical and methodological) links between the calculation of critical loads and integrated assessment modelling.

The material presented here is a summary of the material from CCE Status Reports (see Posch et al. 1995, 1997, 1999). It is complemented by recent EMEP documentation (EMEP 2013; Simpson et al. 2012).

In Section VIII.1 the grid systems used by EMEP are defined, in Section VIII.2 methods for the calculation of percentiles (in one and two dimensions) are presented, and in Section VIII.3 we discuss uses of critical loads in integrated assessment, including different gap closure methods, used to evaluate the differences between a reference scenario (such as past or present situation) and scenarios proposed for future policies.

## VIII.1 GEOGRAPHIC GRID SYSTEMS

To make critical loads usable and useful for the work under the LRTAP Convention, one has to be able to compare them to deposition estimates. Deposition of sulphur and nitrogen compounds have earlier been reported by EMEP on a  $150 \times 150 \text{ km}^2$  grid covering (most of) Europe, then

depositions have become available on a  $50 \times 50 \text{ km}^2$  grid. Both are special cases of the so-called polar stereographic projection, which is described in the following. The latest grid used by EMEP has geographical coordinates (“latitude-longitude grid”, cf. section VIII.1.2).

### VIII.1.1 THE POLAR STEREOGRAPHIC PROJECTION

In the polar stereographic projection each point on the Earth's sphere is projected from the South Pole onto a plane perpendicular to the Earth's axis and intersecting the Earth at a fixed latitude  $\phi_0$  (see Figure A-1 in the CCE Status Report 2001, p. 182.). Consequently, the coordinates  $x$  and  $y$  are obtained from the geographical longitude  $\lambda$  and latitude  $\phi$  (in radians) by the following equations:

(VIII.1)

$$x = x_p + M \cdot \tan\left(\frac{\pi}{4} - \frac{\phi}{2}\right) \cdot \sin(\lambda - \lambda_0)$$

And

(VIII.2)

$$y = y_p - M \cdot \tan\left(\frac{\pi}{4} - \frac{\phi}{2}\right) \cdot \cos(\lambda - \lambda_0)$$

where  $(x_p, y_p)$  are the coordinates of the North Pole;  $\lambda_0$  is a rotation angle, i.e. the longitude parallel to the  $y$ -axis; and  $M$  is the scaling of the  $x$ - $y$  coordinates. In the above definition the  $x$ -values increase and the  $y$ -values decrease when moving towards the equator. For a given unit length (grid size)  $d$  in the  $x$ - $y$  plane the scaling factor  $M$  is given by

$$(VIII.3) \quad M = \frac{R}{d} \cdot (1 + \sin \phi_0)$$

where  $R$  ( $= 6370 \text{ km}$ ) is the radius of the Earth. The inverse transformation, i.e. longitude and latitude as function of  $x$  and  $y$ , is given by

$$(VIII.4) \quad \lambda = \lambda_0 + \arctan\left(\frac{x - x_p}{y_p - y}\right)$$

and

(VIII.5)

$$\phi = \frac{\pi}{2} - 2 \cdot \arctan(r / M)$$

$$\text{with } r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$$

Chapter I

The *arctan* in eq.VIII.5 gives the correct longitude for quadrant 4 ( $x > x_p$  and  $y < y_p$ ) and quadrant 3 ( $x < x_p$  and  $y < y_p$ );  $\pi$  ( $=180^\circ$ ) has to be added for quadrant 1 ( $x > x_p$  and  $y > y_p$ ) and subtracted for quadrant 2 ( $x < x_p$  and  $y > y_p$ ). Note that quadrant 4 is the one covering (most of) Europe.

Chapter II

Every stereographic projection is a so-called conformal projection, i.e. an angle on the sphere remains the same in the projection plane, and vice versa. However, the stereographic projection distorts areas

Chapter III

## VIII.1.2 GRIDS USED FOR CRITICAL LOADS MAPPING UNDER LRTAP CONVENTION

### VIII.1.2.1 THE LATITUDE-LONGITUDE GRID

Chapter IV

The latitude-longitude grid is a plane projection of the EMEP domain based on the World Geodesic System (WGS84, revised in 2004). This grid system is widely used for GPS. This gridding combines information on the shape of the earth, the (nominal) sea level and altitude. In principle, the origin of this gridding system is the mass centre of the earth. However, for the two dimension grid used for mapping critical loads at regional scale is done on a two dimensional grid, the origin of the grid is the intersect between the Greenwich meridian and the equator.

Chapter V

In 2012, at the 36<sup>th</sup> session of the EMEP Steering Body, the EMEP Centres

Chapter VI

Chapter VII

Chapter VIII

Annexes

(even locally), i.e. it is not an equal-area projection (see below).

We define a **grid cell** ( $i, j$ ) as a square in the  $x$ - $y$  plane with side length  $d$  (see eq.8.3) and centre point as the integral part of  $x$  and  $y$ , i.e.

(VIII.6)

$$i = \text{nint}(x) \quad \text{and} \quad j = \text{nint}(y)$$

where 'nint' is the nearest integer (rounding function). Consequently, the four corners of the grid cell have coordinates  $(i \pm 1/2, j \pm 1/2)$ .

suggested to increase spatial resolution of reported emissions. This new grid is defined in a geographic coordinate system (WGS84) with grid cells about 28 km x 28 km. The new domain cover the geographic area between 30°N-82°N latitude and 30°W-90°E longitude (Figure VIII.1, Dore and Vidič, 2012).

This evolution represents a balance between political needs (requests for detailed information at country level), scientific needs (links with biodiversity and climate changes) and technical feasibility (availability of meteorological and emission data at finer scale, computation time) as of 2014 and for the following years.

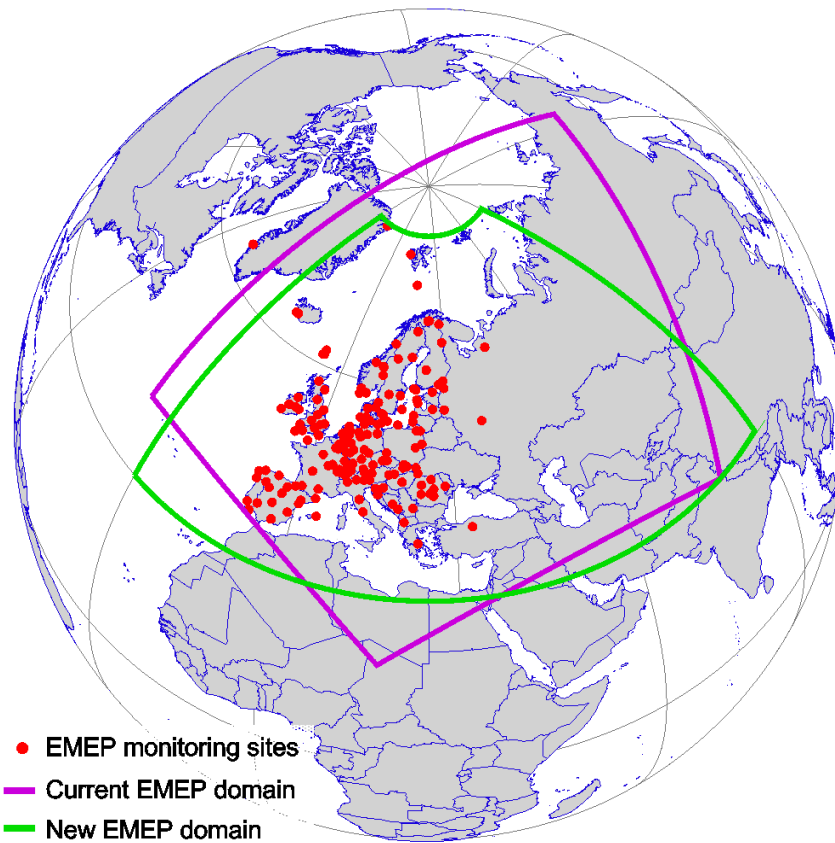


Figure VIII.1 : The EMEP domain over Europe. "Current": Polar-stereographic domain (used since 2008). "New": Latitude-longitude domain (proposed in 2012 (Dore and Vidič, 2012) leading to EB decision 2012/13).

### VIII.1.2.2 THE 50x50 KM<sup>2</sup> GRID (EMEP50 GRID)

The eulerian dispersion model of EMEP/MSC-W produces concentration and deposition fields on a 50 × 50 km<sup>2</sup> grid with the parameters (see also [www.emep.int](http://www.emep.int)):

(VIII.7)

$$d = 50\text{km}, \quad (x_p, y_p) = (8, 110),$$

$$\phi_0 = \frac{\pi}{3} = 60^\circ \text{N}, \quad \lambda_0 = -32^\circ \text{ (i.e. } 32^\circ \text{ W)}$$

yielding  $M=237.7314\dots$

This 50 x 50 km<sup>2</sup> gridding of the EMEP area has been used from 1999 to the beginning of 2013. The domain has been extended from the area covered by the initial 150 x 150 km<sup>2</sup> grid to that presented in Figure VIII.2 in 2008.

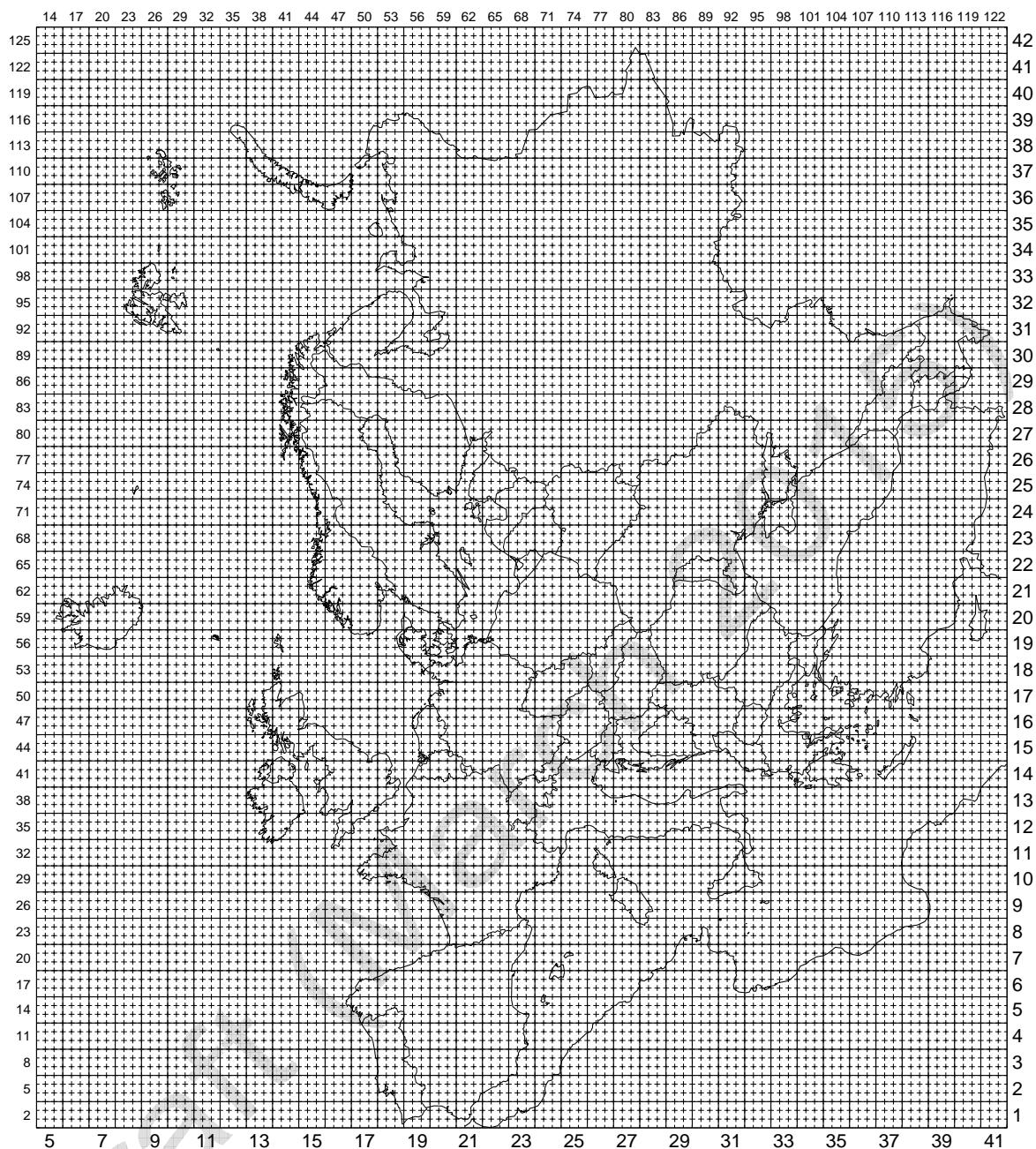


Figure VIII.2 : The EMEP150 grid (solid lines) and EMEP50 grid (dashed lines). The numbers at the bottom and to the right are EMEP150 grid indices; those at the top and to the left are EMEP50 grid indices (every third).

### VIII.1.2.3 THE 150×150 KM<sup>2</sup> GRID (EMEP150 GRID)

The coordinate system used by EMEP/MSC-W for the (old) lagrangian long-range transport model was defined by the following parameters (Saltbones and Dovland 1986):

(VIII.8)

$$d = 150 \text{ km}, \quad (x_p, y_p) = (3, 37),$$

$$\phi_0 = \frac{\pi}{3} = 60^\circ \text{ N}, \quad \lambda_0 = -32^\circ \text{ (i.e. } 32^\circ \text{ W)}$$

which yields  $M=79.2438\dots$



An EMEP150 grid cell ( $i,j$ ) contains  $3 \times 3 = 9$  EMEP50 grid cells ( $m,n$ ) with all combinations of the indices  $m=3i-2, 3i-1, 3i$  and  $n=3j-2, 3j-1, 3j$ . The part of the two

EMEP grid systems covering Europe is also shown in Figure VIII.2.

This gridding of the EMEP area has been used until 1997.

To convert a point ( $xlon,ylat$ ), given in degrees of longitude and latitude, into EMEP coordinates ( $emepi,emepj$ ), the following FORTRAN subroutine can be used:

```

subroutine llemp (xlon,ylat,par,emepi,emepj)
!
! This subroutine computes for a point (xlon,ylat), where xlon is the
! longitude (<0 west of Greenwich) and ylat is the latitude in degrees,
! its EMEP coordinates (emepi,emepj) with parameters given in par().
!
! par(1) ... size of grid cell (km)
! (par(2),par(3)) = (xp,yp) ... EMEP coordinates of the North Pole
!
real(4)          xlon, ylat, par(*), emepi, emepj
!
data Rearth /6370./      ! radius of spherical Earth (km)
data xlon0 /-32./       ! = lambda_0
data drm /1.8660254/    ! = 1+sin(pi/3) = 1+sqrt(3)/2
data pi180 /0.017453293/ ! = pi/180
data pi360 /0.008726646/ ! = pi/360
!
em = (Rearth/par(1))*drm
tp = tan((90.-ylat)*pi360)
rlamp = (xlon-xlon0)*pi180
emepi = par(2)+em*tp*sin(rlamp)
emepj = par(3)-em*tp*cos(rlamp)
!
return
end subroutine llemp

```

EMEP50 coordinates are obtained by calling the above subroutine with  $par(1)=50$ ,  $par(2)=8$  and  $par(3)=110$ ; and EMEP150 coordinates are obtained with  $par(1)=150$ ,  $par(2)=3$  and  $par(3)=37$ . Conversely, the EMEP coordinates of a point can be converted into its longitude and latitude with the following subroutine :

```

subroutine emep11 (emepi,emepj,par,xlon,ylat)
!
! This subroutine computes for a point (emepi,emepj) in the EMEP
! coordinate system, defined by the parameters in par(), its
! longitude xlon and latitude ylat in degrees.
!
! par(1) ... size of grid cell (km)
! (par(2),par(3)) = (xp,yp) ... EMEP coordinates of the North Pole
!
real(4)          emepi, emepj, par(*), xlon, ylat
!
data Rearth /6370./      ! radius of spherical Earth (km)
data xlon0 /-32./       ! = lambda_0
data drm /1.8660254/    ! = 1+sin(pi/3) = 1+sqrt(3)/2
data pi180 /57.2957795/ ! = 180/pi
data pi360 /114.591559/ ! = 360/pi
!
emi = par(1)/(Rearth*drm) ! = 1/M
ex = emepi-par(2)
ey = par(3)-emepj
if (ex == 0. .and. ey == 0.) then ! North Pole
  xlon = xlon0 ! or whatever
else

```

```

      xlon = xlon0+pi180*atan2(ex,ey)
    endif
    r = sqrt(ex*ex+ey*ey)
    ylat = 90.-pi360*atan(r*emi)

                                return

    end subroutine emep11

```

### VIII.1.3 THE AREA OF AN EMEP GRID CELL

As mentioned above, the stereographic projection does not preserve areas, e.g. a  $50 \times 50 \text{ km}^2$  EMEP grid cell is  $2,500 \text{ km}^2$  only in the projection plane, but never on the globe. The area  $A$  of an EMEP grid cell with lower-left corner  $(x_1, y_1)$  and upper-right corner  $(x_2, y_2)$  is given by:

(VIII.9)

$$A(x_1, y_1, x_2, y_2) = 2R^2 \cdot \left\{ \begin{array}{l} I(u_2, v_2) - I(u_1, v_2) \\ -I(u_2, v_1) + I(u_1, v_1) \end{array} \right\}$$

where  $u_1 = (x_1 - x_0)/M$ , etc.; and  $I(u, v)$  is the double integral (see Posch et al. 1997 for details):

$$\begin{aligned}
 I(u, v) &= \iint \frac{2dudv}{(1+u^2+v^2)^2} \\
 &= \frac{v}{\sqrt{1+v^2}} \cdot \arctan \frac{u}{\sqrt{1+v^2}} \\
 &\quad + \frac{u}{\sqrt{1+u^2}} \cdot \arctan \frac{v}{\sqrt{1+u^2}}
 \end{aligned}$$

These two equations allow the calculation of the area of the EMEP grid cell  $(i, j)$  by setting  $(x_1, y_1) = (i-1/2, j-1/2)$  and  $(x_2, y_2) = (i+1/2, j+1/2)$ .

The following FORTRAN functions compute the area of an EMEP grid cell for arbitrary grid indices  $(i, j)$ , for the EMEP50 or the EMEP150 grid, depending on the parameters in  $par()$  (see above):

```

      real function aremep (par,i,j)
!
! Returns the area (in km2) of an ax-parallel cell with
! centerpoint (i,j) in the EMEP grid defined by par().
!
! par(1) ... size of grid cell (km)
! (par(2),par(3)) = (xp,yp) ... EMEP coordinates of the North Pole
!
      integer(4)      i, j
      real(4)         par(*)
!
      external        femep
!
      data Rearth /6370./      ! radius of spherical Earth (km)
      data drm /1.8660254/     ! = 1+sin(pi/3) = 1+sqrt(3)/2
!
      x1 = real(i)-0.5
      y1 = real(j)-0.5
      emi = par(1)/(Rearth*drm) ! = 1/M
      u1 = (x1-par(2))*emi
      v1 = (y1-par(3))*emi
      u2 = u1+emi
      v2 = v1+emi
      ar0 = 2.*Rearth*Rearth
      aremep = ar0*(femep(u2,v2)-femep(u1,v2)-femep(u2,v1)+femep(u1,v1))
                                return
      end function aremep
!
      real function femep (u,v)
!
! Function used in computing the area of an EMEP grid cell.
!
      real(4)         u, v
!
      ui = 1./sqrt(1.+u*u)

```

```

vi = 1./sqrt(1.+v*v)
femep = v*vi*atan(u*vi)+u*ui*atan(v*ui)
return
end function femep

```

The area distortion ratio  $\alpha$ , i.e. the ratio between the area of a small rectangle in the EMEP grid and its corresponding area on the globe is obtained as (Posch et al. 1999):

$$(VIII.11) \quad \alpha = \left( \frac{1 + \sin \phi}{1 + \sin \phi_0} \right)^2$$

which shows that the distortion ratio depends on the latitude  $\phi$  only, and (small) areas are undistorted, i.e.  $\alpha = 1$ , only at  $\phi = \phi_0 = 60^\circ$ .

Draft (March 2015)

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

Annexes

## VIII.2 COMPARING CRITICAL LOADS: CUMULATIVE DISTRIBUTION FUNCTIONS, PERCENTILES AND PROTECTION ISOLINES

Cumulative distribution functions and percentiles allow a statistical description of critical loads and other parameters used by countries.

Cumulative distribution functions are useful to assess and compare the range of critical loads values and of other parameters. Percentiles allow excluding extreme values and therefore to increase the robustness of assessments.

In this section we first define and investigate different methods for calculating percentiles of a cumulative distribution function (cdf) given by a finite number of values. Then we generalise the concept of a percentile to the case in which the cdf is defined by a set of *functions* (critical load functions), resulting in the so-called percentile function (protection isoline).

### VIII.2.1 CUMULATIVE DISTRIBUTION FUNCTION

Assume we have critical load values,  $x_i$ , for  $n$  ecosystems. We sort these values in ascending order, resulting in a sequence  $x_1 \leq x_2 \leq \dots \leq x_n$ . Each value is accompanied by a weight (area)  $A_i$  ( $i=1, \dots, n$ ), characterizing the size (importance) of the respective ecosystem. From these we compute normalized weights  $w_i$  according to

(VIII.12)

$$w_i = A_i / \sum_{j=1}^n A_j, \quad i = 1, \dots, n$$

resulting in:

$$(VIII.13) \quad \sum_{i=1}^n w_i = 1$$

The cumulative distribution function (cdf) of these  $n$  critical load values is then defined by:

(VIII.14)

$$F(x) = \begin{cases} 0 & \text{for } x < x_1 \\ W_k & \text{for } x_k \leq x < x_{k+1} \\ 1 & \text{for } x \geq x_n \end{cases}$$

with

(VIII.15)

$$W_k = \sum_{i=1}^k w_i, \quad k = 1, \dots, n$$

$F(x)$  is the probability of a critical load being smaller than (or equal to)  $x$ , i.e.  $1-F(x)$  is the fraction of ecosystems protected. With this definition  $F(x)$  has the mathematical properties of a cdf:  $F$  is a monotonously increasing right-continuous function with  $F(-\infty)=0$  and  $F(\infty)=1$ . In Figure VIII.3 an example of a cdf is shown; note that the function assumes only a finite number of values.

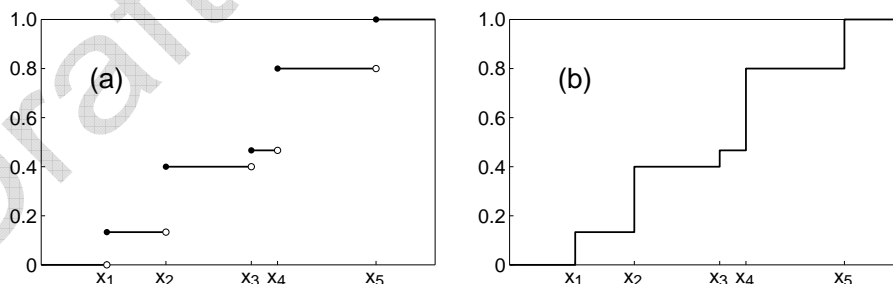


Figure VIII.3 : (a) Example of a cumulative distribution function for  $n=5$  data points ( $x_1 < x_2 < x_3 < x_4 < x_5$ , with weights  $w_1=2/15$ ,  $w_2=4/15$ ,  $w_3=5/15$ ,  $w_4=1/15$ ,  $w_5=3/15$ ). The filled (empty) circles indicate whether a point is part (not part) of the function. (b) The same cdf is drawn by connecting all points, the way a cdf is usually displayed.

## VIII.2.2 QUANTILES AND PERCENTILES

All ecosystems in a region (grid cell) are protected, if deposition stays below the smallest critical load values. However, to discard outliers and to account for uncertainties in the critical load calculations, but also to ensure that a sufficient percentage of ecosystems are protected, (low) percentiles of the cdf are compared to the deposition.

The  $q$ -th **quantile** ( $0 \leq q \leq 1$ ) of a cdf  $F$ , denoted by  $x_q$ , is the value satisfying

$$(VIII.16) \quad F(x_q) = q$$

which means that  $x_q$ , viewed as a function of  $q$ , is the inverse of the cdf, i.e.  $x_q = F^{-1}(q)$ .

**Percentiles** are obtained by scaling quantiles to 100, i.e. the  $p$ -th percentile is the  $(p/100)$ -th quantile. Other terms used are *median* for the 50-th percentile, lower and upper *quartile* for the 25-th and 75-th percentile, respectively. Note that the  $p$ -th percentile critical load protects  $100-p$  percent of the ecosystems.

Computing quantiles, i.e. the inverse of a cdf given by a finite number of points poses a problem: due to the discrete nature of the cdf, a unique inverse simply does not exist. For many values of  $q$  no value  $x_q$  exists at

all so that eq.VIII.16 holds; and for the  $n$  values  $x_i$  such a value exists (i.e.  $q = F(x_i)$ ), but the resulting quantile is not unique – every value between  $x_i$  and  $x_{i+1}$  could be taken (see Figure VIII.2). Therefore, the cdf is approximated (interpolated) by a function which allows solving eq.VIII.16 for every  $q$ . There is neither a unique approximation, nor is there a single accepted way for calculating percentiles: Posch et al. (1993) discuss six methods for calculating percentiles. Note that commonly definitions are given for data with identical weights (i.e.  $w_i = 1/n$ ), but the generalization to arbitrary weights is mostly straightforward. It should be also born in mind that the differences between different approximation methods vanish when the number of points becomes very large (and all weights small).

In the following we have a closer look at two types of quantile functions: (a) those derived from linearly interpolating the cdf, and (b) those using the empirical cdf. After defining their equations for arbitrary weights we discuss their advantages and disadvantages.

### VIII.2.2.1 LINEAR INTERPOLATION OF THE CDF

In this case the quantile function is the inverse of the linearly interpolated cdf given by:

$$(VIII.17) \quad x_q = \begin{cases} x_1 & \text{for } q \leq w_1 = W_1 \\ x_k + (x_{k+1} - x_k) \cdot \frac{q - W_k}{w_{k+1}} & \text{for } W_k < q \leq W_{k+1} \\ & k = 1, \dots, n \end{cases}$$

where the  $W_k$  are defined in eq.VIII.15. An example is shown in Figure VIII.4a.

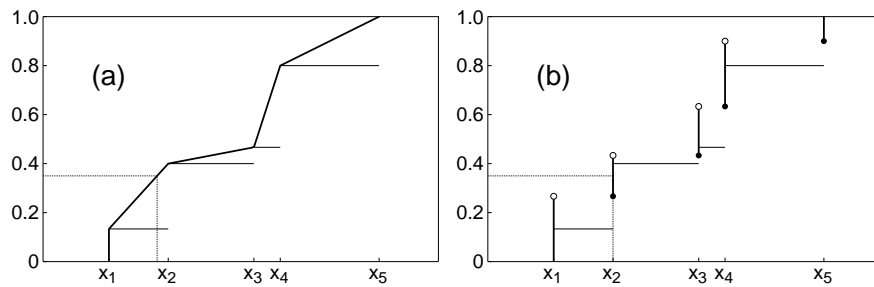


Figure VIII.4 : Examples of the two quantile functions discussed in the text. Values and weights are the same as in Figure VIII.4. The filled (empty) circles indicate whether a point is part (not part) of the function. The thin horizontal lines indicate the cumulative distribution function. Note that for almost all values of  $q$  (e.g.  $q=0.35$ ) the resulting quantile is smaller in (a) than in (b).

The advantage of this quantile function is that it is continuous, i.e. a small change in  $q$  leads to only a small change in the resulting quantile  $x_q$ . However, it has the following three disadvantages:

(i) In case of two (or more) identical data points the definition of the quantile function is not unique: for identical critical load values the shape of the interpolation function depends on the order of the

weights (see Figures VIII.5 a,a'). This could be resolved by sorting the weights of identical data points according to size (smallest first, as in Figures VIII.5a.b). This minimizes the difference to the empirical distribution function (see below), but requires fairly complicated (and time-consuming) routines for the actual computations.

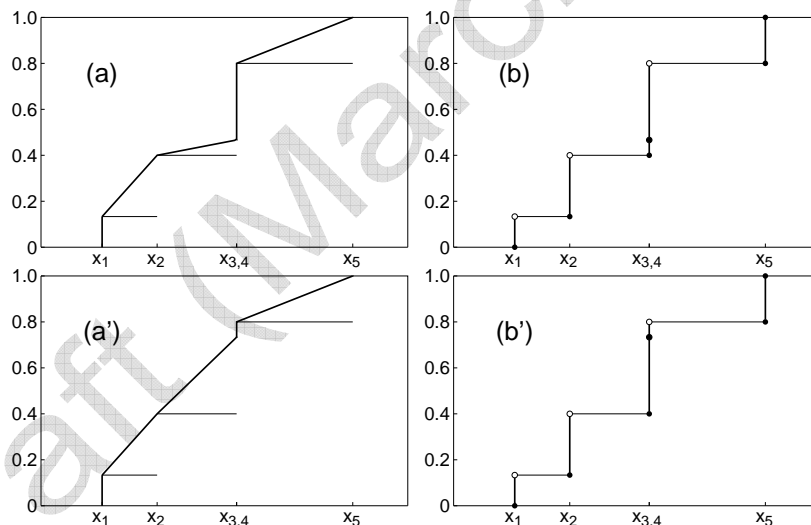


Figure VIII.5 : Examples of the two quantile functions discussed in the text. Values and weights are the same as in Figure VIII.2, except that  $x_3=x_4$  (compare Figure VIII.5). Note, that for the linearly interpolated quantile function (a,a') its shape depends on the order of the weights for the identical values.

(ii) As mentioned above, a critical load  $x_q$  is selected to protect the  $(1-q)$ -th fraction of the ecosystems within a given region (grid cell). However, for the linear interpolated quantile function certain choices of  $q$  result in  $x_q$ -values which are *below* the actual value needed to protect a fraction  $1-q$  of the ecosystems (see example in Figure VIII.4). This is protective for the ecosystems, but may lead to higher costs for abatement.

(iii) The computation of quantiles is not order-preserving when using linear interpolation. We say the order is preserved by a quantile function, if the following holds for two cdfs:

(VIII.18)

$$F_1(x) \leq F_2(x) \text{ for all } x \Rightarrow x_q^{(1)} \leq x_q^{(2)} \text{ for all } q$$

i.e. the smaller cdf leads to smaller quantiles. In Figure 8.5a an example is shown with two data sets for the same  $n$  ecosystems,  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  with common weights  $w_1, \dots, w_n$  and the property  $x_i < y_i$  for  $i=1, \dots, n$  (e.g.  $CL_{min}$ 's and  $CL_{max}$ 's). But for certain values of  $q$  it turns out that  $x_q > y_q$  when computed by linear interpolation (Fig. VIII.6a).

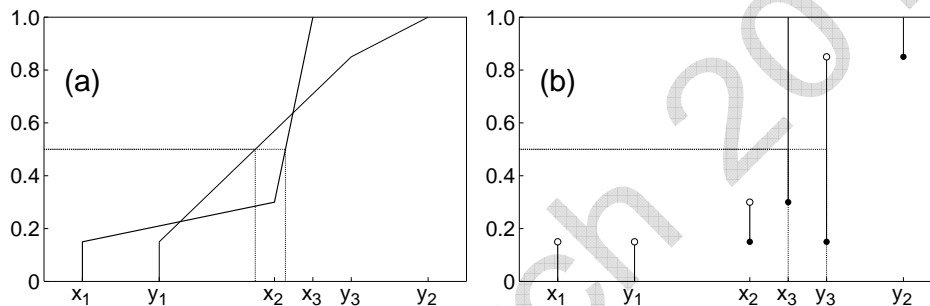


Figure VIII.6 : Example of two quantile functions for 3 values each ( $x_1, x_2, x_3$  and  $y_1, y_2, y_3$ ) and common weights  $w_1, w_2, w_3$  and the property  $x_i < y_i$  for  $i=1,2,3$ . However, in case (a) the median  $x_{0.5}$  is greater than the median  $y_{0.5}$ .

### VIII.2.2.2 EMPIRICAL DISTRIBUTION FUNCTION:

In this case the quantile function assumes only values defining the cdf:

(VIII.19)

$$x_q = \begin{cases} x_1 & \text{for } q < w_1 = W_1 \\ x_k & \text{for } W_{k-1} \leq q < W_k, k = 2, \dots, n-1 \\ x_n & \text{for } q \geq W_{n-1} \end{cases}$$

An example of this quantile function is shown in Figure VIII.5b. The disadvantage of this quantile function is that it is not continuous, i.e. a very small change in  $q$  may lead to a significant change in the quantile  $x_q$  (jump from  $x_i$  to  $x_{i+1}$ ).

However, none of the disadvantages of the linear interpolation holds for this function, but:

- (i) identical values do not lead to ambiguities (see Figures VIII.5b,b'),
- (ii) the quantile  $x_q$  protects (at least) a fraction  $q$  of the ecosystems (see Figure VIII.4b), and
- (iii) the computation of quantiles is order-preserving (see eq.VIII.18 and Figure VIII.6b).

It is especially property (iii) which makes the empirical distribution function the only viable choice for computing percentiles. The following FORTRAN subroutine computes the  $q$ -quantile of a given vector of data with a corresponding vector of weights. The data have to be sorted in ascending order, but the weights do not have to be normalised to one.

```

subroutine qantilcw (q,num,vec,wei,xq)
!
! This subroutine computes the q-quantile xq of the num values in vec()
! - sorted in ascending order - with corresponding weights wei()
! from the empirical distribution function.
!
integer(4)      num
real(4)        q, vec(*), wei(*), xq
!
if (num == 0) stop 'Quantile of nothing?!'
if (q < 0. .or. q > 1.) stop 'q outside [0,1]!'
!
wsum = wei(1)
do k = 2,num
  wsum = wsum+wei(k)
  if (vec(k) < vec(k-1)) stop 'Data not sorted!'
end do
!
qw = q*wsum
sum = 0.
do k = 1,num
  sum = sum+wei(k)
  if (qw < sum) then
    xq = vec(k)
    return
  end if
end do
xq = vec(num) ! if q=1
return
end subroutine qantilcw

```

### VIII.2.3 PERCENTILE FUNCTIONS AND PROTECTION ISOLINES

In this section we generalize of the concept of cumulative distribution function (cdf) and quantile (percentile) to the case when the data (e.g. critical loads) are given as a functions (rather than as single values), which is the case when considering two pollutants (e.g. sulphur and nitrogen in the case of acidification), leading to the so-called percentile function or (ecosystem) protection isoline.

In the following we assume that a (critical load) function is defined by a set of pairs of values (nodes)  $(x_j, y_j)$ ,  $(j=1, \dots, m)$ , and the function is given by connecting  $(x_1, y_1)$  with  $(x_2, y_2)$  etc., in this way generating a polygon in the x-y plane. We denote this polygon by:

(VIII.21)

$$f = [(x_1, y_1), \dots, (x_m, y_m)]$$

For the values  $x_j$  and  $y_j$  we assume that:

(VIII.22)

$$0 = x_1 \leq x_2 \leq \dots \leq x_m \quad \text{and} \quad y_1 \geq y_2 \geq \dots \geq y_m = 0$$

i.e. the nodes on the polygon are numbered from left to right, starting on the y-axis and ending on the x-axis. Eq.VIII.22 also ensures that the polygon is monotonically decreasing, when considered as a function of  $x$  or  $y$ . (Alternatively, the numbering could start on the x-axis, etc.). With the notation  $(x, y) < f$  we mean that the point  $(x, y)$  lies below the polygon (i.e. critical loads are not exceeded).

Considering the critical load for S and N acidity the critical load function for an ecosystem is defined by 3 values, namely  $CL_{min}(N)$ ,  $CL_{max}(N)$  and  $CL_{max}(S)$ , and as a polygon with  $m=3$  nodes it is written according to eq.VIII.21 as:

(VIII.23)

$$CLF = [(0, CL_{max}(S)), (CL_{min}(N), CL_{max}(S)), (CL_{max}(N), 0)]$$

where we assumed that the N-deposition is plotted along the x-axis and the S-deposition along the y-axis.



Now we assume that we have  $n$  critical load functions  $f_1, \dots, f_n$  with respective weights  $w_1, \dots, w_n$  ( $\sum w_i = 1$ ). In general it will not be possible to sort these critical load functions, i.e. it is not possible to say that  $f_i$  is larger or smaller than  $f_j$ , because  $CL_{max}(S)$  for  $f_i$  could be larger and  $CL_{max}(N)$  smaller than the corresponding values for  $f_j$  (see Figure VIII.7 for examples). Nevertheless, we can define a cumulative distribution function  $F$  in the following way:

$$(VIII.24) \quad F(x, y) = \sum_{(x,y) < f_i} w_i$$

meaning that for a given point  $(x, y)$  we sum all weights  $w_i$  for which  $(x, y) < f_i$ , i.e. for which there is no exceedance. Obviously  $0 \leq F(x, y) \leq 1$ , and  $F$  has also otherwise all properties of a (two-dimensional) cdf. A percentile  $p$  is now easily defined as the intersection of such a function with a horizontal plane at height  $q = p/100$ . The result (projected onto the  $x$ - $y$  plane) is a curve, more precisely a polygon which has the property defined in eq.VIII.22. Let  $f_q$  be the quantile (percentile) function for a given  $q$ , then every point  $(x, y)$ , i.e. every pair of  $N$  and  $S$  deposition, with  $(x, y) < f_q$  protects (at

least) a fraction of  $1 - q$  of the ecosystems; and  $f_q$  is also called a (ecosystem) protection isoline. Note that protection isolines for the same set of polygons (critical load functions) do not intersect (although they might partly coincide), and for  $r < s$   $f_r$  lies below  $f_s$ .

Since an exact computation of a percentile function is hardly feasible (especially in case of a large number of critical load functions), we have to use an approximate method (see Figure VIII.7): we draw rays through the origin of the  $x$ - $y$  plane (i.e. lines with a constant  $S:N$  deposition ratio) and compute the intersections of these rays with all critical load functions (small circles in Figure VIII.7a). For each ray the intersection points are sorted according to their distance from the origin and the chosen quantiles of these distances are calculated according to eq.VIII.19. Finally, the resulting quantile values are connected to obtain the percentile functions (protection isolines). Obviously, the more rays are used in this procedure the more accurate are the protection isolines. As Figure VIII.8b shows, a protection isoline need not be convex.

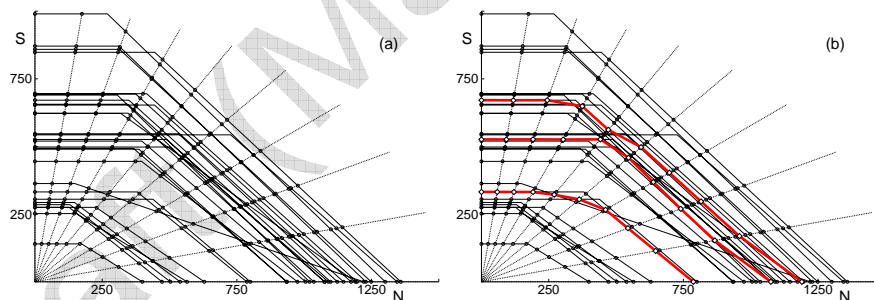


Figure VIII.7 : Computation of protection isolines: (a) set of critical load functions and intersection of these CL-functions with rays from the origin (small circles); (b) computing the percentiles ( $q=0.25, 0.50$  and  $0.75$  in this case) along each ray (small diamonds) and connecting them to obtain the protection isolines (thick [red] lines).

## VIII.3 CRITICAL LOAD EXCEEDANCES USED IN INTEGRATED ASSESSMENT MODELLING

### VIII.3.1 GAP CLOSURE METHODS

Except for the earliest protocols, integrated assessment modellers have used uniform *percentage* reductions of the excess deposition (so-called *gap closures*) to define emission reduction scenarios. In the following we summarize the different gap closure methods used and illustrate them for the case of a single pollutant. This section follows largely Posch et al. (2001).

In the 1994 Sulphur Protocol, only sulphur was considered as acidifying pollutant (N deposition was fixed; it determined, together with N uptake and immobilization, the sulphur fraction). Furthermore, taking into account the uncertainties in the CL calculations, it was decided to use the 5-th percentile of the critical load cdf in a grid cell as the only value representing the ecosystem sensitivity of that cell. And *the* exceedance was simply the difference between the (current) S deposition and that 5-th percentile critical load. This is illustrated in Figure VIII.8a): Critical loads

and deposition are plotted along the horizontal axis and the (relative) ecosystem area along the vertical axis. The thick solid and the thick broken lines are two examples of critical load cdfs (which have the same 5-th percentile critical load, indicated by 'CL'). 'D0' indicates the (present) deposition, which is higher than the CLs for 85% of the ecosystem area. The difference between 'D0' and 'CL' is the exceedance in that grid cell. It was decided to reduce the exceedance everywhere by a fixed percentage, i.e. to 'close the gap' between (present) deposition and (5-th percentile) critical load. In Figure VIII.8a, a *deposition gap closure* of 60% is shown as an example. As can be seen, a fixed deposition gap closure can result in very different improvements in ecosystem protection percentages (55% vs. 22%), depending on the shape of the critical load cdf.

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

Annexes

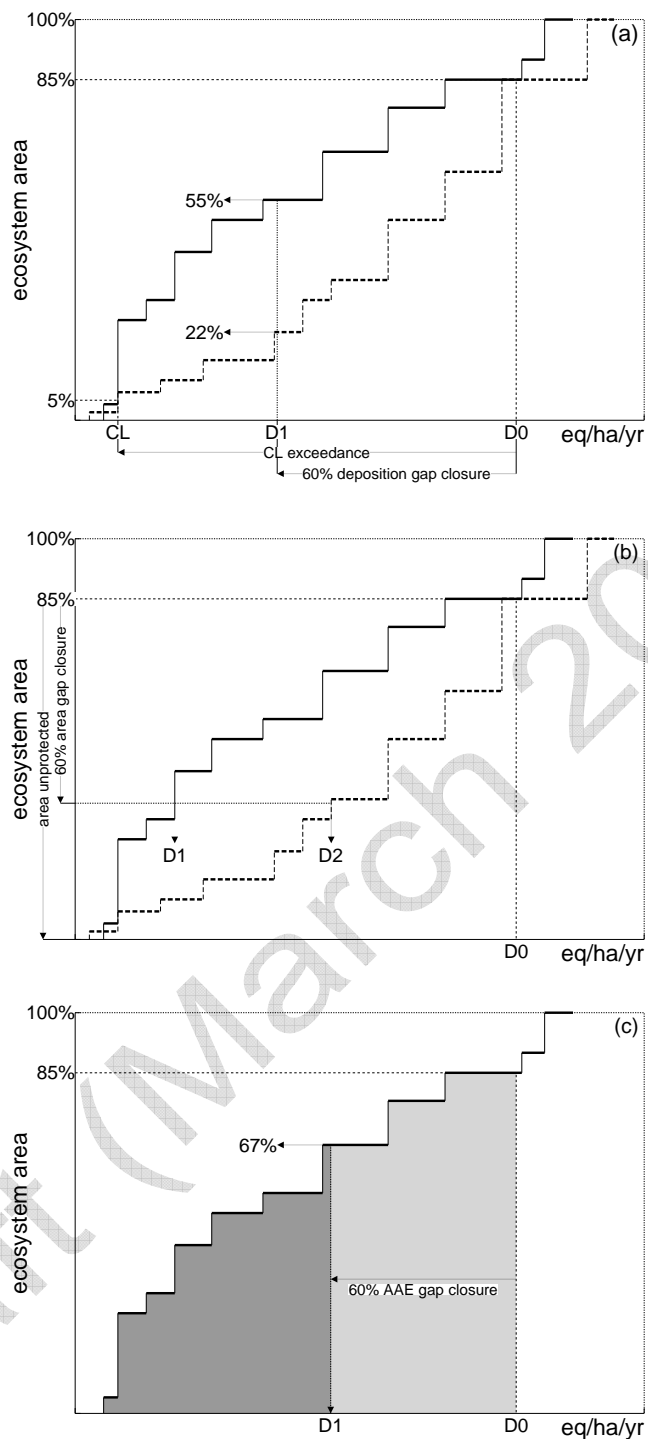


Figure VIII.8 : Cumulative distribution function (thick solid line) of critical loads and different methods of gap closure: (a) deposition gap closure, (b) ecosystem gap closure, and (c) accumulated exceedance (AE) gap closure. The thick dashed line in (a) and (b) depict another cdf, illustrating how different ecosystem protection follows from the same deposition gap closure (a), or how different deposition reductions are required to achieve the same protection level (b).

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

Annexes

To take into account all critical loads within a grid cell (and not only the 5-th percentile), it was suggested to use an *ecosystem area gap closure* instead of the deposition gap closure. This is illustrated in Figure VIII.8b: for a given deposition 'D0' the ecosystem area unprotected, i.e. with deposition exceeding the critical loads can be read from the vertical axis. After agreeing to a certain (percent) reduction of the unprotected area (e.g. 60%), it is easy to compute for a given cdf the required deposition reduction ('D1' and 'D2' in Figure VIII.8b). Another important reason to use the ecosystem area gap closure is that it can be easily generalized to two (or more) pollutants, which is not the case for a deposition-based exceedance. This generalization became necessary for the negotiations of the 1999 Gothenburg Protocol, as both N and S contribute to acidification. Critical load values have been replaced by critical load functions and percentiles replaced by ecosystem protection isolines (see above). However, the use of the area gap closure becomes problematic if only a few critical load values or functions are given for a grid cell. In such a case the cdf becomes highly

discontinuous, and small changes in deposition may result in either no increase in the protected area at all or large jumps in the area protected.

To remedy the problem with the area gap closure caused by discontinuous cdfs, the accumulated exceedance (AE) concept has been introduced (see above). In the case of one pollutant, the AE is given as the area under the cdf of the critical loads (the entire grey-shaded area in Figure VIII.8c). Deposition reductions are now negotiated in terms of an AE (or AAE) *gap closure*, also illustrated in Figure VIII.8c: a 60% AE gap closure is achieved by a deposition 'D1' which reduces the total grey area by 60%, resulting in the dark grey area; also the corresponding protection percentage (67%) can be easily derived. The greatest advantage of the AE and AAE is that it varies smoothly as deposition is varied, even for highly discontinuous cdfs, thus facilitating optimization calculations in integrated assessment. The advantages and disadvantages of the three gap closure methods described above are summarized in the following table.

	Advantages	Disadvantages
<b>Deposition gap closure</b> (used for the 1994 Sulphur Protocol)	<ul style="list-style-type: none"> <li>• Easy to use even for discontinuous cdfs (e.g. grid cells with only one CL).</li> </ul>	<ul style="list-style-type: none"> <li>• Takes only one CL value (e.g. 5th percentile) into account.</li> <li>• May result in no increase in protected area.</li> <li>• Difficult to define for two pollutants.</li> </ul>
<b>Ecosystem area gap closure</b> (used for the EU Acidification Strategy)	<ul style="list-style-type: none"> <li>• In line with the goals of CL use (maximum ecosystem protection).</li> <li>• Easy to apply to any number of pollutants.</li> </ul>	<ul style="list-style-type: none"> <li>• Difficult (or even impossible) to define a gap closure for discontinuous cdfs (e.g. grid cells with only one CL).</li> </ul>
<b>Accumulated Exceedance (AE) gap closure</b> (used for the 1999 Gothenburg Protocol)	<ul style="list-style-type: none"> <li>• AE (and AAE) is a smooth and convex function of deposition even for discontinuous cdfs.</li> </ul>	<ul style="list-style-type: none"> <li>• AE stretches the limits of the critical load definition.*</li> <li>• Exceedance definition not unique for two or more pollutants.</li> </ul>

\* It assumes a linear damage function. However, this feature could also be an advantage.

### VIII.3.2 LINEAR EMISSION-EXCEEDANCE RELATIONSHIPS

The change to higher resolutions of the EMEP grid and the introduction of new pollutants (in particular PM) has led to a significant increase of required calculations for optimization exercises of scenarios. In order to keep computing times at a workable level, the methodology to calculate critical loads exceedances has been simplified. The new approach has been inspired by the Life Cycle Impact Analysis (LCIA) community which uses the simplest approach possible, i.e. a linear relationship between emission (changes) and impact (changes) on ecosystems. Models and factors are described and defined in Posch et al. (2005).

In practice, linear relationships are defined between the emission and the average accumulated exceedances for each country. This linearization requires deposition fields computed exactly with a full atmospheric transport model for the

reference scenario (for instance “current legislation” scenario). The coefficients of the linear relationship are “impact” or “damage” factors. They are site or country dependant. The impact factors are computed by changing the emission of one pollutant at the time in a given source region, compared to the reference scenario, leaving the emissions of the other pollutants and all other source regions unchanged (further details in Posch et al. 2005).

The approach makes it feasible to assess several policy scenarios. Errors on the results remain small if the assessed scenario remains close to the reference scenario: tests have shown that when the emissions are reduced by up to 20%, the linear model produces good approximations. Besides, approximations become poor when exceedances become close to 0.

### VIII.4 REFERENCES

- Dore CJ, Vidič S (2012) Considerations of changing the EMEP Grid. Document prepared by the EMEP Centres MSC-W, MSC-E, CEIP and CIAM, EMEP, Geneva, Switzerland. 9 pp.
- EMEP (2013) Transboundary Acidification, Eutrophication and Ground Level Ozone in Europe in 2011. Norwegian Meteorological Institute, Oslo, Norway, 205 pp.
- Posch M, Kämäri J, Johansson M, Forsius M (1993) Displaying inter- and intra-regional variability of large-scale survey results. *Environmetrics* 4: 341-352
- Posch M, De Smet PAM, Hettelingh J-P, Downing RJ (eds) (1995) Calculation and mapping of critical thresholds in Europe. Status Report 1995, Coordination Centre for Effects, RIVM Report 259101004, Bilthoven, Netherlands, iv+198 pp. [www.wge-cce.org](http://www.wge-cce.org)
- Posch M, Hettelingh J-P, De Smet PAM, Downing RJ (eds) (1997) Calculation and mapping of critical thresholds in Europe. Status Report 1997, Coordination Centre for Effects, RIVM Report 259101007, Bilthoven, Netherlands, iv+163 pp. [www.wge-cce.org](http://www.wge-cce.org)
- Posch M, De Smet PAM, Hettelingh J-P, Downing RJ (eds) (1999) Calculation and mapping of critical thresholds in Europe. Status Report 1999, Coordination Centre for Effects, RIVM Report 259101009, Bilthoven, Netherlands, iv+165 pp. [www.wge-cce.org](http://www.wge-cce.org)
- Posch M, Hettelingh J-P, De Smet PAM (2001) Characterization of critical load exceedances in Europe. *Water, Air and Soil Pollution* 130: 1139-1144

Chapter I

Posch M, Hettelingh J-P, Heyes C (2005) Use of critical loads in integrated assessment modelling. In: M Posch, J Slootweg, J-P Hettelingh (eds), European Critical loads and dynamic modelling: CCE Status Report 2005. RIVM Report259101016, Coordination Centre for Effects, Bilthoven, The Netherlands, pp. 71-76

Chapter II

Saltbones J, Dovland H (1986) Emissions of sulphur dioxide in Europe in 1980 and 1983. EMEP/CCC Report 1/86, Norwegian Institute for Air Research, Lillestrøm (now in Kjeller), Norway

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

**Chapter VIII**

Annexes

Simpson D, Benedictow A, Berge H, Bergström R, Emberson LD, Fagerli H, Flechard CR, Hayman GD, Gauss M, Jonson JE, Jenkin ME, Nyiri A, Richter C, Semeena VS, Tsyro S, Tuovinen JP, Valdebenito A, Wind P (2012) The EMEP MSC-W chemical transport model - technical description. *Atmospheric Chemistry and Physics* 12(16): 7825-7865.

Draft (March 2015)