

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Sensitivity Analysis of the 2-dimensional application
of the Generic Ecological Model to the North Sea

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Katarzyna Salacinska

Delft, the Netherlands
August 2008

MSc THESIS APPLIED MATHEMATICS

“Sensitivity Analysis of the 2-dimensional application of the Generic Ecological Model to the North Sea”

Katarzyna Salacinska

**Delft University of Technology
Deltares – Dutch Institute for Delta Technology**



Daily supervisor

Dr. Ghada El Serafy, Deltares

Responsible professor

Prof. Dr. Ir. Arnold Heemink, Delft
Institute of Applied Mathematics

Other thesis committee members:

Dr. Dorota Kurowicka, TU Delft
Dr. Remus Hanea, TU Delft

August, 2008

Delft, the Netherlands

Acknowledgements

“At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.”

Albert Schweitzer

I would like to give my thanks to Dr. Ghada El Serafy, Prof. Arnold Heemink and Dr. Remus Hanea for supervising my work and providing me with very helpful comments and corrections. Ghada, thank you very much for your patience and motivation. Every day spent at Deltares allowed me to gain useful experience and brought me a lot of joy. Moreover, I would like to thank Hans Los who kindly familiarized me with the model and Anouk Blauw for her contribution to the project.

I am grateful to Prof. Roger Cooke and Prof. Jolanta Misiewicz for the given opportunity to participate in the “Risk and Environmental Modelling” Msc program. The bridge between TU Delft and the university of Zielona Gora should always be open. Crossing it allows students to make their dreams come true. Even when it was windy, when it was rainy, during exam periods, or at the moments of yearning for home, I have never regretted my decision to come here. What I have learnt at this university is a priceless gift.

I wish to express deep gratitude to Dr. Dorota Kurowicka for her help, support and last but not least her time. Thank you for sharing your knowledge and experience with me.

My parents, my sister, my friends, people who never let me down, thank you very much for the inspiration and enthusiasm. You gave me faith because you believed in me. You made me reach the place where I am now and I hope one day I will pay my debt of gratitude.

Table of contents:

1	Introduction.....	9
2	Model description - the GEM model.....	10
3	Overview of methods for sensitivity analysis	20
3.1	The Morris method	23
3.2	Factorial Designs	25
3.3	Variance-based methods	26
3.4	Metamodelling.....	28
4	Analysis of the results.....	29
4.1	The Morris method - maximal chlorophyll-a concentration.	35
4.2	The Morris method - annual average chlorophyll-a concentration.	46
4.3	Variance-based analysis - maximal chlorophyll-a concentration.....	54
4.4	Variance-based analysis - annual average chlorophyll-a concentration.	62
5	Discussion	66
6	References.....	69
7	Appendix.....	70

1 Introduction

Control and reduction of undesirable ecosystem changes caused mainly by growing human demand, anthropogenic impacts and climate change are considered as main items in the international legislator's precedence. Phytoplankton is an important indicator for water quality both in fresh and salt water, as it affects many factors related to the ecological quality of the water, such as turbidity, oxygen depletion and productivity of the system. Due to the fact that the harmful algae might become a plague and cause damage to tourism, mussel industry and farmers, it is essential to be able to make accurate predictions about algae's future composition and abundance as well as when and where algal blooms could occur. Forecasting enables managers not only to keep a control over unwanted changes, but also to evaluate potential effects of some management strategies. As a consequence, the development of the Generic Ecological Model for estuarine and costal waters has been initiated [2]. It consists of separate physical, chemical and ecological model components (waves, morphology, hydrodynamics, suspended sediments) which are then coupled together to build one generic and flexible modelling tool. The validation results demonstrated accuracy of the GEM model for various key parameters both in spatial and temporal dimensions for a variety of different water systems without the need for continuous remaking of new models for each different system, site or objective, or for major re-parametrisation [1]. Since chlorophyll-a is the pigment essential for photosynthesis and is commonly found in most types of phytoplankton, its concentration is used to determine the concentration of phytoplankton in a given water system. In view of the fact that satellite observations of the chlorophyll-a concentration for the North Sea are available, it is sensible to take advantage of them and improve the forecasts of the model with use of data assimilation techniques, within this particular application, in order to obtain truthful, accurate and realistic estimations. To do so, a sensitivity analysis might be seen as a helpful tool in understanding the role of uncertainty and complexity in the model as well as in reducing dimensions considered for the data assimilation procedure. Many possible methods are available for conducting sensitivity analysis. The choice of the proper ones depends on such considerations as the number of input factors, possible interactions among them and finally it is also limited by the computational cost of running the model. In this report the results of sensitivity analysis performed for the 2D GEM application to the North Sea will be presented.

The report starts with a short introduction to the model of interest and its general principle. Further, in Section 4 a verification of the possible methods for sensitivity analysis of this model is illustrated. Section 5 presents a selection of appropriate ones and the obtained results. Finally, some remarks on the applied methodology and further use of produced results are made and recommendations for future work are given.

2 Model description - the GEM model

Over the past decades, a relatively large number of models have been developed for the simulation of nutrient cycles, primary producers and ecosystem functioning. They all differ markedly in model complexity expressed in terms of description of the water quality and ecological processes, area included, and level of temporal and spatial resolution. Most phytoplankton models solve a set of differential equations, in which the growth of each species is expressed as the product of several terms based upon the availability of resources. However, this approach is not very well suited to describe competition between a relatively large number of phytoplankton species since competition between them becomes a fairly complicated function of limiting and non-limiting resources especially when it is combined with a great number of taxonomic groups. GEM integrates the best aspects of other models, it includes both physical, chemical and ecological processes at a sufficient level of detail and in the consistent way.

GEM is part of the Delft 3D [3] integrated modeling system of Deltares, former WL|Delft Hydraulics, which includes separate modules for hydrodynamics as well as for waves, morphology and suspended sediments. In order to apply the model to a specific area one has to define, besides the ecological processes and parameter settings, the input for schematisation and transport, loadings, boundaries, forcings and initial conditions.

Schematisation

A model application requires a hydrodynamic calculation which is then coupled to the water quality – ecological modeling instrument BLOOM/GEM. It is allowed to use different time steps and grids for different processes, for example transport and water quality processes, that will result in the reduction of simulation time. It is also possible to model with curvilinear and variable grids, which can be better adapted to spatial gradients using fine resolution where necessary and a much coarser resolution elsewhere. The application to the southern North Sea is given below:

aggregated grid

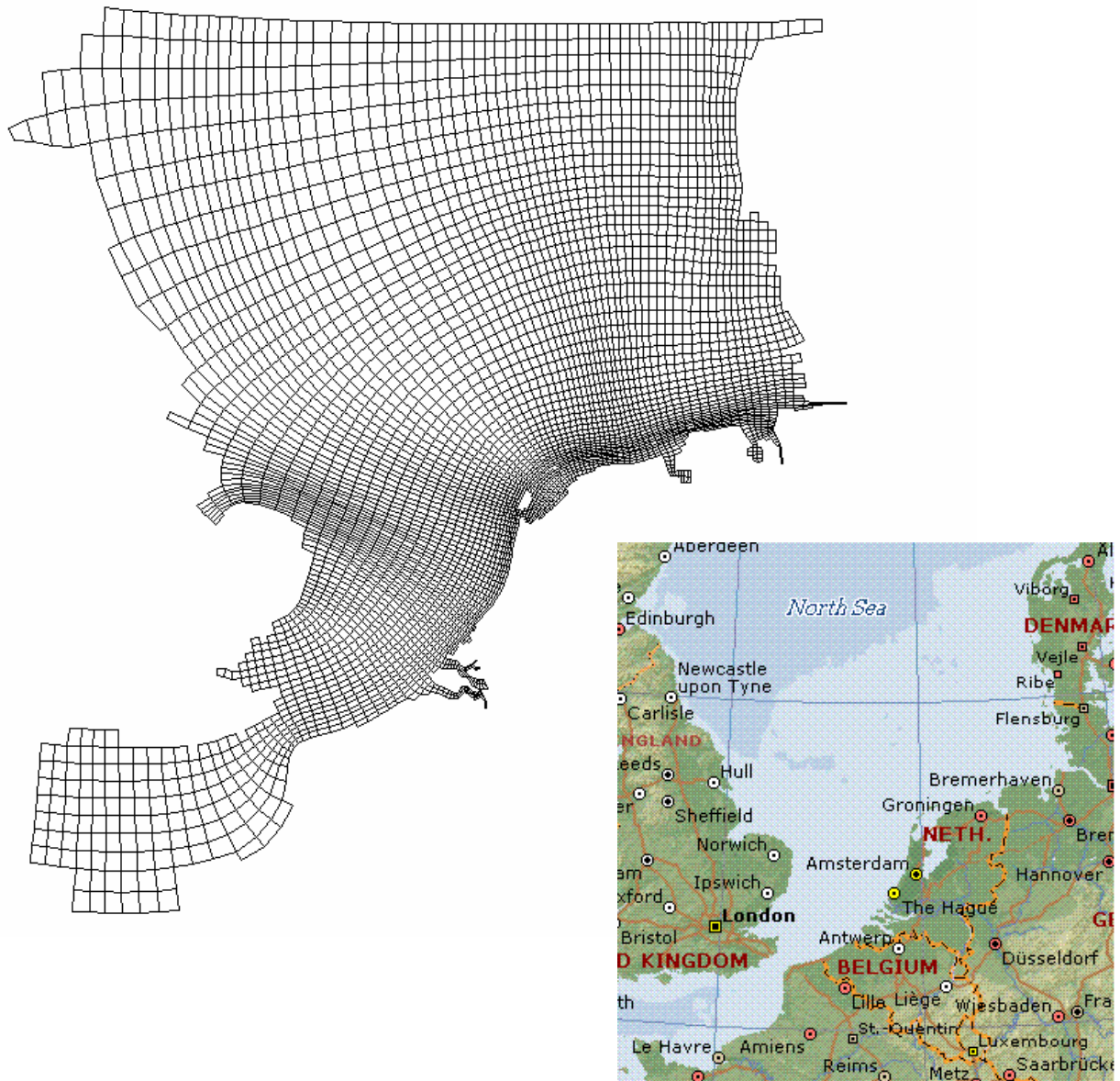


Figure 3.1 Zuno grid

Its horizontal resolution is relatively high in the costal areas of interest, notably the Dutch costal zone (approximately 1x1 km) and coarser in the northern part of the area included in the model (approximately 20x20 km). The grid size is 134 x 65 that gives in total 8710 of grid points. The boundaries are defined along the land, Atlantic Ocean interface (57 degrees N) and English Channel.

Hydrodynamics

GEM can be combined with any hydrodynamic model. For 2D mode the hydrodynamic simulations are executed in 3D mode and then vertically averaged. Transport of substances is described by the advection-dispersion equation:

$$(1) \quad \frac{\partial C}{\partial t} = -\frac{\partial uC}{\partial x} - \frac{\partial vC}{\partial y} - \frac{\partial wC}{\partial z} + \frac{\partial}{\partial x}(D_x \frac{\partial C}{\partial x}) + \frac{\partial}{\partial y}(D_y \frac{\partial C}{\partial y}) + \frac{\partial}{\partial z}(D_z \frac{\partial C}{\partial z}) + S(x, y, z)$$

where

C – concentration [ML^{-3}],

(u,v,w) – velocity vector [LT^{-1}],

(D_x, D_y, D_z) – components of the dispersion tensor [L^2T^{-1}],

x,y,z – coordinates in three spatial dimensions [L],

S – source or sink of mass due to physical, chemical and biological processes [$\text{ML}^{-3}\text{T}^{-1}$].

Equation 1 states that the change of the concentration in time is caused by advective transport due to translation with the velocity vector (u,v,w) and by dispersive transport, plus addition or extraction of mass (sink/source). The source/sink term represents waste loads as well as various water quality and ecological processes. A wide range of numerical schemes is available to solve the transport part in the advection-dispersion equation. In the Delft 3D suit user may choose the cyclic method or Van-Leer 2 scheme, which are both finite difference methods. The hydrodynamic conditions (velocities, water elevations, density, salinity, vertical eddy viscosity and vertical eddy diffusivity) calculated in the Delft3D-FLOW module are used as the input to DELWAQ, the program for modelling water quality and aquatic ecology.

State variables

GEM considers three nutrient cycles, namely nitrogen, phosphorus and silicate and this composition is fully sufficient to determine the water quality in case of four phytoplankton species: Diatoms, Flagellates, Dino-Flagellates and Phaeocystis. The carbon cycle is partially modelled, and a mass-balance of organic carbon is made. The nutrient cycle has three major pools: dissolved inorganic nutrients, living organic matter and dead organic matter. Consequently, the following state variables are included in the model:

Dissolved inorganic state:

- NO_3 – representing the sum of nitrate and nitrite,
- NH_4 – representing ammonia,
- PO_4 – representing orthophosphate,
- Si – representing silicate.

Dead particulate organic matter:

- POC – representing particulate organic carbon,
- POP – representing particulate organic phosphorus,
- PON - representing particulate organic nitrogen,
- POSi – representing opal silicate.

Organic matter in the sediment:

- POCS
- POPS
- PONS
- POSiS

Additional model variables are:

- dissolved oxygen
- salinity
- zooplankton biomass

Within each of the species groups, three phenotypes are defined regarding the adaptation to different environmental conditions. A suitability of a type is determined by the ratio of its requirements and its growth rate. Hence the following types can be distinguished:

- Energy types – high growth rate, low mortality rates, high nitrogen to carbon (N:C) and phosphorus to carbon (P:C) ratio,
- Nitrogen types – lower N:C ratio, lower maximum growth rates, higher mortality rates, higher settling velocities and higher chlorophyll content,
- Phosphorus types – lower P:C ratio, lower maximum growth rates, higher mortality rates, lower settling velocities and lower chlorophyll content.

This contributes to another twelve state variables.

Moreover, one can define the following objects among the substances:

- Continuity - a special type of conservative tracer, it has no physical or chemical meaning but it is used to establish the numerical correctness and stability of the simulation,
- Fraction fresh water from constant discharge,
- Fraction fresh water from variable discharge¹.

Modelled processes

One or more state variables of the model might appear, disappear or change into another state variable due to some physical, biological or chemical reactions. The processes they are involved in are the primary production (making of organic compounds from carbon dioxide through photosynthesis), respiration and mortality (phytoplankton biomass is released partially as dead particulate matter and partially as inorganic nutrients via autolysis). The change in the water conditions causes a shift in species composition. As it was mentioned before, the model follows the principle of competition between groups of algae. It was shown mathematically that selecting the combination of species groups that uses the limiting factor (nutrient or light) most efficiently basically means maximizing the total net production of the phytoplankton community. Therefore, the linear programming technique is used to determine the species composition [1]. The optimization algorithm selects the resource that is most likely to become limiting and the best adapted type under prevailing conditions:

Optimization: maximize the net growth

Constraints:

1. The biomass increase of any of the species groups cannot exceed the maximum net growth rate (production minus respiration) at actual temperature and light intensity.
2. The mortality rate of any of the species groups cannot exceed the maximum mortality rate at actual temperature and salinity.
3. The total extinction of light by phytoplankton cannot exceed the threshold level where the light intensity becomes insufficient to maintain further net growth.

¹ For 2D GEM the hydrodynamic simulation of one spring-neap cycle, that is about 14 days, is used and then the same cycle is repeated at the end of this period until the end of the simulation. Hence, variable flows cannot be determined in the hydrodynamic model and the yearly averaged flows are used instead (FrCon). GEM produces variable discharges every 10-day period (FrVar). If the difference between the actual and average fresh water discharge is too large, at a particular location and time, the transport based upon the average hydrodynamic computation is not accurate. By comparing this actual fresh water discharge with the average value used by hydrodynamics, one can approximate the actual salinity within GEM. In case of models with the hydrodynamic forcing throughout the year, therefore in agreement with the ecological part, the fresh water discharges are the same. Thus two separate tracers for water discharge would not be needed.

4. The total uptake of each of the nutrients (N, P, Si) must not exceed the availability. The total available amount of a nutrient is defined as the sum of dissolved inorganic nutrient plus the amount of the nutrients in phytoplankton.

Subsequently, the algorithm considers the next potentially limiting factor and again selects the best adapted phytoplankton type.

The optimisation technique finds the new biomass of each algal phenotype at the end of a time step. Later, they are summed up to compute the biomass of each species. The rates of growth, production, mortality and autolysis are derived from the change of the algae biomasses over a time step.

Since primary production is strongly influenced by light availability, the correct calculation of light conditions in the water column is essential. Extinction of light by substances is modelled as an exponential decrease of light intensity with depth according to the Lambert-Beer formula:

$$(2) \quad I_z = I_0 * e^{-kz}$$

where

z – depth [m],

I_z - underwater light intensity at z [Wm^{-2}],

I_0 - surface irradiance [Wm^{-2}],

k – extinction coefficient [m^{-1}].

The extinction coefficient is calculated as the sum of the extinction by inorganic suspended matter, particulate organic matter (POM), phytoplankton, dissolved humic substances and background extinction.

Algae mortality produces detritus. Mineralization of detritus, in the water column as well as in the bottom sediment, produces inorganic nutrients. In the water column, the decomposition rate is dependent on the nutrient stoichiometry, it is high for high nutrient content in detritus. Particulate organic matter in the sediment is the result of settling of phytoplankton and dead particulate matter from the water column. For each phytoplankton species a separate settling velocity, constant in time, is specified. In 3D applications the effect of turbulences is included additionally expressed by vertical dispersion between the water layers. For POM only one settling velocity is determined. The decomposition rate of POM in the sediment depends only

on the temperature. Remineralised nutrients are released back into the overlying water column. Some part of the particulate organic matter in the sediment is essentially removed from the model as the result of burial. The burial rate is the calibration parameter and it is a constant fraction of the annual amount of particulate organic matter in the sediment.

Optionally resuspension and the effect of bottom shear stress can be included.

GEM solves the mass balance for oxygen, in which several oxygen producing and oxygen consuming processes are considered. Oxygen is produced by algae (primary production) and is consumed by algal respiration, by mineralization of detritus and by nitrification, since nitrification and denitrification are the first order processes modelled in GEM. Exchange of oxygen with the atmosphere (reaeration) can result in either a gain or loss of oxygen in the water column. The reaeration rate in most GEM applications is a function of wind speed and water depth [1].

The description of the grazing pressure includes the uptake of phytoplankton biomass and detritus for food, the production, respiration and mortality of biomass, and the excretion of detritus and nutrients. The biomass of grazers is imposed as a forcing and then the model simulates its effect. One of attractive features of the GEM model is that the biomass of primary consumers is corrected during simulation if the food shortage occurs what indicates the first step in dynamic modelling of the grazing process. It is allowed to define up to five types of grazers, which may be species groups or individual species of zooplankton and zoobenthos. Consumed food is either assimilated as grazer biomass, respired or egested as detritus [3 p.8-52]. The grazing module is included in the application to the North Sea and grazers biomass is one of the state variables.

In the figure below all possible processes are shown, but a decision which ones to include in the model must be made on the basis of the objective of the application, i.e. fresh or coastal water system.

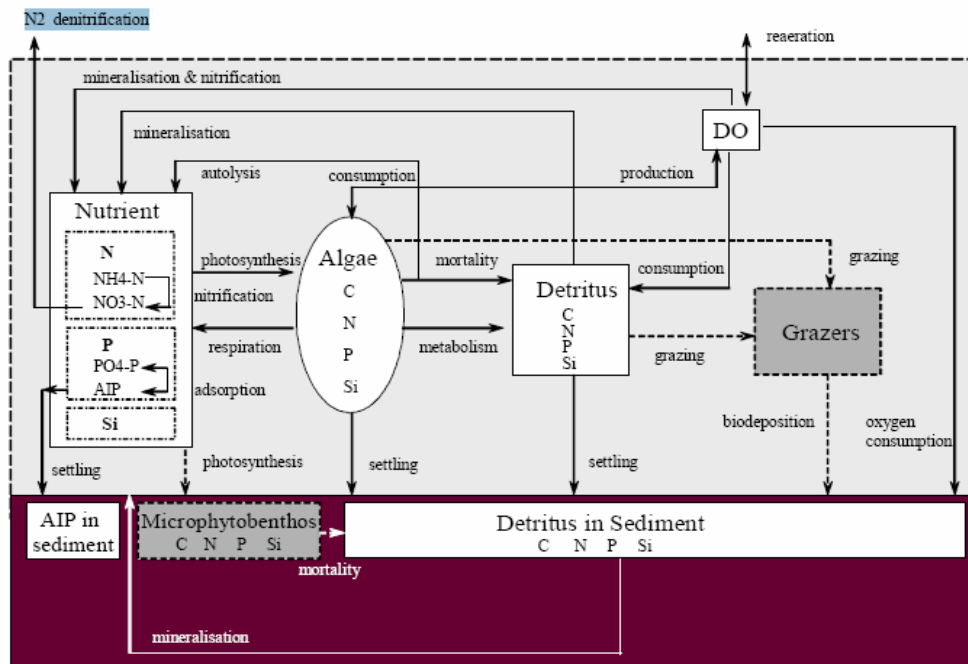


Figure 3.2

Time step

The primary production and nutrient model uses the same vertical and horizontal grid as the hydrodynamic model. For the North Sea a time step of 30 minutes is used for the substance transport. Primary production and oxygen dynamics are also simulated at a 30 minute time step, all other biological and chemical processes are simulated using a 24 hours time step.

Boundary conditions

The model adopts the exchange of water masses on the boundaries from the hydrodynamic simulations. In the North Sea application boundaries concentrations are time dependent. Data for the English Channel boundary and for the northern boundary concentrations are specified as a monthly time series. Moreover the model set-up considers 36 point sources and sinks of nutrients and fresh water from the main Dutch, German, French and UK rivers.

Default parameter setting has been calibrated for the North Sea and has proven to be applicable for a range of other coastal ecosystems.

Validation

GEM has been successfully applied for over a decade in a range of different consultancies and studies by Deltares that have formed the basis for several major policy and management decisions, regarding also the infrastructural development. Some examples of the ecosystems for which it was used, except the southern North Sea, are

- Sea of Marmara – a deep, stratified coastal sea in Turkey,
- Veerse Meer – a coastal, saline lake in the southwestern part of the Netherlands,
- Venice Lagoon – very shallow, saline estuarine lagoon system bordering the Italian city of Venice along the Mediterranean coast.

More information about the above applications can be found in [1].

Model validation is one of the most important steps in the model building sequence. The testing procedure is carried out before the model is accepted and used to support decision making. Precisely, a model validation is a confirmation that a model, within its domain of applicability, possesses a satisfactory accuracy consistent with the intended application of the model.

GEM has been validated extensively both in its 2D and more recently in its 3D version with respect to its main outputs (dissolved nutrients, oxygen, chlorophyll, extinction coefficient, species groups). The goodness of fit has been determined by use of the cost function:

$$(3) \quad C_x = \frac{\sum |M_{x,t} - D_{x,t}| / n}{sd_x} * ((1 - c) + c(1 - r_x))$$

where C_x is the normalized deviation per station, annual value, $M_{x,t}$ is mean value of the model results per station per month, $D_{x,t}$ is mean value of the in situ data per station per month, sd_x is standard deviation of the annual mean based on the monthly means of the in situ data (df=11), n is 12 months, c is 0.5 and r_x is the correlation over time between $M_{x,t}$ and $D_{x,t}$.

The validation results (presented in [1] and [4]) demonstrated accuracy of the GEM model for various key parameters both in spatial and temporal dimensions for a variety of different water systems without the need for continuous remaking of new models for each different system, site or objective, or for major reparametrisation. In most cases the values of the cost function were low, namely values between 0 and 2 were obtained, that according to the ratings criteria proposed by of Radach and Moll indicate that the model produces consistently good and acceptable results:

Rating	Condition
Very good	$0 < cf \leq 1$
Good	$1 < cf \leq 2$
Reasonable	$2 < cf \leq 3$
Poor	$3 < cf \leq$

Table 3.1 Ratings criteria for the cost functions

Only in shallow, stratified, dynamic coastal areas for example the Wadden Sea, the model performance was poor, what was the consequence of model simplification done by not

including the sediment as a separate layer in the model grid. However, these limitations can be overcome by extra modifications to a specific model application. They are currently being discussed and further improvements to GEM are going to be made.

Below, the validation result for 3D GEM simulation for the North Sea at station Schouwen10, located 10 km offshore in the Delta region of the Netherlands, are presented:

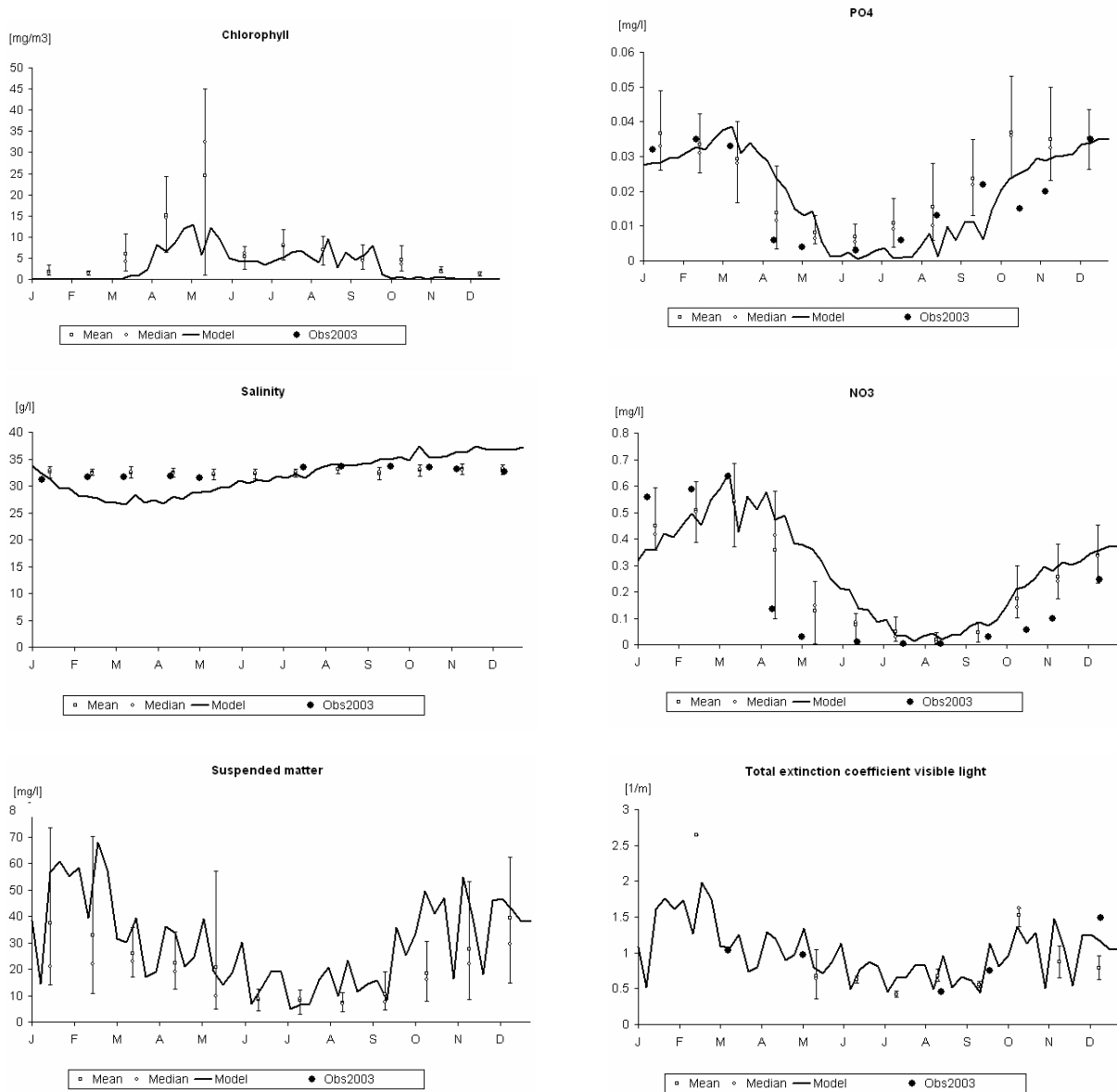


Figure 3.3 Validation result for 2D GEM simulation at station Schouwen10 (Chlorophyll-a concentration, PO₄, salinity, NO₃, suspended matter and total extinction coefficient). Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996-2002

One can see that model results correspond very well with the measurements.

3 Overview of methods for sensitivity analysis

The possible definition of sensitivity analysis that appears most often in the literature is the following: the study of how the uncertainty in the model output can be apportioned to different sources of uncertainty in the model input. In other words, it allows to determine which of the input parameters are more important in influencing the uncertainty of the model output.

Sensitivity analysis can be of use in the growing field of numerical simulation, where mathematical and computational models are used for the study of systems, especially complex ones. Conducting sensitivity analysis might be indeed useful, in particular to uncover technical limitations in the model, identify critical regions in the space of the inputs and to establish priorities for research.

Below, the general concept of sensitivity analysis is presented. Specifications for the test case are given at the beginning of Section 5.

The following steps in sensitivity analysis are distinguished [7]:

- 1. Establish the goal of the analysis and consequently define the form of the output function.**

A poor definition of the objective of the sensitivity analysis can lead to confused or inconclusive results. Diverse statistical tests and measures may be thrown at the problem, producing the range of different factor rankings and giving no clue which one to believe or privilege. To avoid this situation, one should define a relevant importance of a factor for the exercise in question. It is called a “setting”. The possible settings are:

- The Factor Prioritization (FP):

Used to identify a group of factors that, when fixed to their true values, lead to the greatest reduction in the variance of the output, i.e. account for most of the output variance. Later, one can decide to rank them as the object of further analysis or measurement.

- The Factor Fixing (FF):

Used to identify a group of factors that, left free to vary over their range of uncertainty, make no significant contribution to the variance of the output, i.e identify non-influential factors.

In case of very complex models, computationally expensive to evaluate, that involve a large number of factors, it is right to come up with a list of truly important factors among many potentially important ones. Remaining parameters can be fixed anywhere in their range of variation without appreciably affecting a specific output of interest.

- The Variance Cutting (VC):

Used for the reduction of the output variance below a given tolerance. It is done by fixing the smallest number of input factor.

- The Factor Mapping (FM):

Used for study on which values of the input factors lead to model realization in a given range of the output space.

Moreover, exploring the sensitivity of several model outputs seems to be beneficial, however one should focus on the key inference suggested by the model while presenting the obtained results in order to provide an auditor with simplicity of understanding.

2. Decide on which input factors should be included in the analysis.

One should be as careful and objective as possible in deciding on the input for sensitivity analysis. Clearly, the more variables are promoted to the rank of input and allowed to vary the greater the variance to be expected in the model prediction. This can lead to the conclusion that the model prediction varies so widely as to be of no practical use. From the perspective of sensitivity analysis, even for a moderately large number of factors, the input space can become too large to explore it thoroughly. Some methods can fail because for a great number of parameters they cannot embrace all possible factor-level combinations. To avoid these difficulties, it is desirable to detect at an early stage of the experimentation which factors are important and which are not. The unimportant factors are then dropped from consideration by fixing them at some reasonable values, and further experimentation will only be carried out with the more important factors.

3. Choose a distribution function for each of the input factors.

To conduct sensitivity analysis the uncertainty in the inputs must be known to ascertain how a given model is influenced by it.

4. Choose a sensitivity analysis method:

Several different approaches have been proposed for sensitivity analysis, due to the intrinsic difficulty of building an effective measure over a space of variation for the inputs, consideration of its structure and the model behaviour. Therefore the following issues should be taken into account when deciding on the appropriate method to use:

- The question that should be answered – quantitative or qualitative?

- The number of model evaluations that has to be run and the time needed to run one simulation.
- The presence of dependence structure between the input factors.

Having all mentioned key issues thought-out one is ready to apply the selected methods and analyze the obtained results.

In practice two different schools of thought may be identified, the local sensitivity analysis and the global one. The first one investigates the local response of the output obtained by varying factors one at a time while holding the others fixed to a base value. This approach involves partial derivatives:

$$(4) \quad \frac{\partial Y}{\partial X_i}$$

or its normalized version:

$$(5) \quad \frac{\sigma_{X_i}}{\sigma_Y} \frac{\partial Y}{\partial X_i}$$

and it is mostly met in the literature. Here X_i and Y denote an input and an output of interest respectively.

The method is attractive with respect to efficiency in computer time, in particular when the incremental ratios are used as a sensitive measure. The model has to be executed only a few times compared to the dimension of the array of derivatives to be computed. On the other hand, the fatal limitation is that it is unwarranted when the model input is uncertain and when the model is non-linear. In other words, derivatives are only informative at the base point where they are computed and do not provide an exploration of the rest of the space of the input factors. Hence, only one point of the factors' space is explored. Usually, the base point is a point of maximum probability in the set where the model response takes a specific value. More about the local methods can be found in [11].

The second school is more ambitious in two respects: firstly the space of the inputs is explored within a finite or even infinite region and secondly the variation of the input induced by a factor is taken globally – that is averaged over the variation of all the factors. A few well-known global approaches will be described further in this report due to the mentioned limitations of the local approach. Methods, that are not able to deal with a great number of parameters (more than 20) and rather long simulation time will be passed over, for the sake of

specification to the analysis of GEM. Methods, that were found the most suitable for the problem at hand will be described in details.

3.1 The Morris method

Screening designs are a convenient choice when the problem setting is FF. These designs are conceived to deal with models containing tens or hundreds of input parameters efficiently. As a drawback, they tend to provide qualitative sensitivity measures – they rank the input factors in order of importance but do not quantify how much more influential a given factor is than another. Since the amount of information revealed via a sensitivity analysis depends heavily on the number of sample points that are simulated and on where they are located, an attractive feature of the Morris method is that it attempts to explore several regions of the input space therefore it can be regarded as global. Moreover it is model independent – it does not require any model assumptions to be applied. It is a tool to determine which factors may be considered to have effects which are:

- Negligible
- Linear or additive
- Non-linear or involved in interactions with other factors

To illustrate it, assume that a k-dimensional vector of input factors: X_1, \dots, X_k is given, each of which varies across p selected levels in [0,1] (if a factor follows the distribution other than uniform it is opportune to select the levels in the space of quantiles of the distribution). For a given value $\mathbf{x}=(x_1, \dots, x_k)$ the elementary effect of the input factor X_i (d_i) is defined as follows:

$$(6) \quad d_i(\mathbf{x}) = \frac{y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(\mathbf{x})}{\Delta}$$

where $\Delta \in \{\frac{1}{p-1}, \dots, 1 - \frac{1}{p-1}\}$ is fixed and y denotes a model response [7].

The distribution F_i of elementary effects associated with the i th input factor is obtained by randomly sampling different \mathbf{x} from the input space. The sensitivity measure μ and σ are respectively the estimates of the mean and standard deviation of the distribution F_i (it is recommended to consider the mean μ^* of the distribution of absolute values of the elementary effects to deal with effects of different signs). Generally, μ estimates the overall effect of the

factor on the output and σ estimates the ensemble of the second and higher-order effects that the factor is involved.

In order to build an efficient sample to estimate the elementary effects we need to generate r trajectories each of them of $(k+1)$ sampling points with the key property that two consecutive points differ in only one component and the first point has been selected at least once to be increased by Δ . Each trajectory allows the computation of an elementary effect for each factor $i, i=1, \dots, k$. Once r elementary effects per input are available the statistics μ_i, μ_i^* and σ_i can be computed by using the same estimators that would be used with independent random samples:

$$(7) \quad \mu_i = \frac{1}{r} \sum_{j=1}^r d_i^j$$

$$(8) \quad \mu_i^* = \frac{1}{r} \sum_{j=1}^r |d_i^j|$$

$$(9) \quad \sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^r (d_i^j - \mu)^2$$

A graphical representation in the (μ, σ) plane allows for a better interpretation of results since it takes into account two sensitivity measures at the same time.

It is known that the standard error of the mean of a sample \bar{x} is given by the formula:

$$(10) \quad SEM_{\bar{x}} = \frac{STD_{\bar{x}}}{\sqrt{n}}$$

where n is a sample size and STD is the standard deviation of the population. In practice, because the standard deviation of the population is usually unknown, the standard deviation of the sample is used instead. Moreover the α % confidence interval for the estimated mean is written as

$$(11) \quad EM_{\bar{x}} = \pm Z_{\frac{\alpha}{2}} SEM_{\bar{x}}$$

where $Z_{\frac{\alpha}{2}}$ denotes the $\frac{\alpha}{2}$ % quantile of the standard normal distribution.

Hence, in our case for 10 different trajectories and $\alpha = 5$ % the formula (23) becomes:

$$(12) \quad EM_{\bar{x}} = \pm 2 \frac{STD_{\bar{x}}}{\sqrt{10}}$$

Two lines $STD_{\bar{x}} = \frac{\sqrt{10}}{2} EM_{\bar{x}}$ and $STD_{\bar{x}} = -\frac{\sqrt{10}}{2} EM_{\bar{x}}$ will be also graphed to make the analysis clearer. If the point (μ_i, σ_i) lies outside the wedge formed by these two lines this may suggest significant evidence that the expectation of the elementary effect is not zero.

The total cost of applying the Morris method is $r(k+1)$ model evaluations hence it is a relatively cheap method. A critical choice related to the implementation of the method is the choice of the parameters p and Δ . The first one is strictly related to the choice of r . When a value of p is high it must be coupled with a high value of r to avoid a greater number of unexplored levels.

It can be concluded that the Morris method is effective, conceptually simple and it can be thought of as an expansion of a derivative-based approach: μ and σ are the mean and standard deviation of approximation of derivatives at different points of the input space. However, it overcomes the limitation mentioned in the Section 4.1. When a single trajectory is considered and the variations of input factors are small, it reduces to an incremental ratio estimation. Moreover, it has a number of advantages with respect to other screening methods that are widely applied in the literature.

It is worth mentioning that if X_i follows some non-uniform distribution G_i with a finite domain different than $[0,1]$ or even infinite domain we can normalize it by the following transformation:

$$Y_i = G_i(X_i)$$

and then the domain of Y_i is the interval $[0,1]$ by the definition of cumulative distribution function. Given a normalized parameter value Y_i between 0 and 1 an analyst can determine the associated value for X_i through the inverse transformation.

3.2 Factorial Designs

Full factorial designs, in the most common case, consider k factors each of them with only two levels, in particular max and min, and this procedure provides the smallest number of runs when many factors have to be investigated. Then the model results for all possible combinations of high and low values, that represent the corners of the corresponding k -dimensional hypercube, are generated. The main effect is computed as the difference between the average response for the high level and average response for the low level. The drawback is that the full factorial design requires then 2^k experiments to approximate $2^k - 1$ effects that are k

main effects, $\frac{k!}{2!(k-2)!}$ two-factor interactions effects, ..., one k-factor interaction effect.

However, it is possible to reduce the cost by picking out the fraction such as 1/2, 1/4, etc. This technique is called the fractional factorial design. However, it results in confounding effects of different orders, thus the information about the important high-order effects can be lost [12].

It is a useful tool in case of dealing with the FF setting, although in the analysis of GEM it was decided against applying this method for the sake of the Morris approach. The latter gives more or less the same amount of information about the influence of input factors on the output, but with a smaller number of model runs comparing to the fractional factorial design, that must have been used for GEM, that involves a great number of factors.

3.3 Variance-based methods

They are linked to both the FP and FF setting.

Interesting features are:

- Model independence
- Capacity to capture the influence of the full range of variation of each input factor
- Appreciation of interaction effects among input parameters

The drawback of the variance-based methods is their computational cost.

The procedure is the following. Let factor X_i be fixed at a particular value x_i^* . One can compute the variance of the output $V_{X_{\neq i}}(Y | X_i = x_i^*)$ taken over all factors but X_i ($X_{\neq i}$). In order to make it independent of the value of x_i^* it is advisable to take the mean over all possible values of X_i that is $E_{X_i}(V_{X_{\neq i}}(Y | X_i = x_i^*))$. Finally, the incorporation of the fact that $V(Y) = E(V(X)) + V(E(X))$ and division of the result by unconditional variance of Y yield to the first-order sensitivity index of X_i on Y:

$$(13) \quad S_i = \frac{V_{X_i}(E_{X_{\neq i}}(Y | X_i))}{V(Y)}$$

that represents the main effect contribution of the given factor to the variance of the output Y. A high value indicates an important variable, moreover it is always between 0 and 1.

Computing S_i may be tricky, as it involves a conditional expectation which is not generally available in closed form. For the estimation of the conditional expectation different methods are in use for instance: the ‘Pedestrian’ method, Kernel estimation methods or optimization techniques.

Interactions represent important features of models and are more difficult to detect than first-order effects. Two factors are said to interact when their effect on Y cannot be expressed as a sum of their single effects thus by the sum of the first-order sensitivity indices. The joint effect of the pair (X_i, X_j) can be measured by $V(E(Y | X_i, X_j))$.

Let us denote by V_{ij} the second-order effect:

$$(14) \quad V_{ij} = V(E(Y | X_i, X_j)) - V(E(Y | X_i)) - V(E(Y | X_j))$$

Analogous formulas can be written for higher-order terms. Moreover, $V_i = V(E(Y | X_i))$.

Then applying ANOVA-HDMR decomposition and assuming that factors are independent:

$$(15) \quad V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots k}$$

Dividing both sides of (15) by $V(Y)$ the following result is obtained:

$$(16) \quad \sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \dots + S_{12\dots k} = 1$$

This means that even for non-additive models it is possible to recover 100 % of the variance of Y [6]. Variance-based methods provide a theoretical framework whereby - provided one has the patience to compute all the interaction terms-one can achieve a full understanding of the model’s sensitivity pattern.

The number of terms in above equation increases exponentially with the number of input factors $(2^k - 1)$.

Instead of such an exhausting computation it is reasonable to use another approach namely the total effect indices. The total effect accounts for the total contribution to the output variation due to factor X_i (first order effect plus all higher-order effects due to interactions).

For a model with three factors, for example, the total effect of X_1 is the sum of all the terms that included it:

$$(17) \quad S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$$

To find the total effect of X_i one can simply decompose the output variance $V(Y)$ in terms of main effect and residual conditioning this time with respect to all the factors but one:

$$(18) \quad V(Y) = E(V(Y | X_{\neq i})) + V(E(Y | X_{\neq i}))$$

Then the total effect index for X_i can be computed as:

$$(19) \quad S_{T_i} = \frac{E(V(Y | X_{\neq i}))}{V(Y)} = 1 - \frac{V(E(Y | X_{\neq i}))}{V(Y)}$$

To estimate it using the Monte-Carlo approach $N(k+2)$, with $N=1000$ or higher, model evaluations are needed. Using the Random Balance Designs this number can be reduced to N (first-order terms only).

Usually, the set of all S_i together with the set of all S_{T_i} is computed to obtain a fairly good description of the model sensitivities at a reasonable cost. The condition $S_{T_i} = 0$ is necessary and sufficient for X_i to be a non-influential factor. If $S_{T_i} \cong 0$, then X_i can be fixed at any value within its range of uncertainty without appreciably affecting the value of the output variance $V(Y)$.

The Morris method is seen as another method that can be used as an excellent and cheap proxy for the total sensitivity indices (19).

3.4 Metamodelling

In the literature on sensitivity analysis there has been a growing interest in metamodelling and smoothing techniques. The first one is an important and powerful method that allows replacement of the original complex model with a cheaper one, which is operationally equivalent. More precisely it tries to answer the question: if we were to approximate function f such that $Y=f(X_1, \dots, X_k)$ that is usually unknown to the analyst with a function of one single parameter X_i what function $g_i^*(X_i)$ would produce minimum loss?

Metamodelling is based on the kernel regression that under certain regularity conditions consistently approximate g^* (the best estimate is obviously $E(Y|X_i)$). Then, the quantity

$$\frac{V(g_i^*(X_i))}{V(Y)} = \text{corr}(g_i^*(X_i), Y)$$

is used to provide the fraction of the variability of Y that is explained with the best predictor based on X_i . Hence, it is clear how metamodelling is linked with the theory of variance-based sensitivity analysis. The information about the available methods for approximating function f and further references can be found in [6]. The drawback is that it is very difficult to compute in case of very complex models. In addition, the total effect estimation is a weak element of this technique since it requires adding all the first-order and interactions terms associated with each input factor. Moreover, adopting this method

requires a great data resource in order to carry out goodness-of-fit tests for the obtained estimate. These limitations drew back metamodelling from the further consideration.

4 Analysis of the results

For the analysis of GEM the Factor Fixing was chosen from the list of settings in Section 4 - the nature of the model corresponds well with the characteristics mentioned in the description of this setting. After an extensive literature study on the global methods for sensitivity analysis and careful investigation of their applications it was decided to use the Morris method together with quantitative methods based on the variance. The Morris method required the implementation and it was done in Matlab. For the latter Unicorn, the software, open source, developed at the Department of Mathematics of Delft University of Technology [5], turned out to be a very helpful tool providing its user with such capabilities as:

- estimating the correlation ratio (in particular the first-order sensitivity index) by using a polynomial fit procedure to the conditional expectation,
- calculating simple statistics (mean, variance) of a single input factor,
- calculating product moment correlation, rank correlation, regression coefficient, partial correlation and partial regression coefficient of a pair of input variables,
- calculating multiple regression coefficient for a set of input parameters,
- many attractive plotting options i.e. scatterplots.

Justification of this choice follows directly from the known model behavior, namely its complexity, non-linearity, engagement of a great number of parameters, small amount of gathered data and, the most importantly, a relatively long time of a single simulation.

In total the model makes use of 406 parameters, corresponding to the processes dedicated to modelling the nutrient cycling and primary production, but focusing on the chlorophyll-a concentration the group of 71 has been selected, on the basis of experience, that is seen as the one containing potentially the most influential factors and it was further processed to avoid difficulties in conducting sensitivity analysis for the entire number of factors. The list of parameters within the group together with justification of the choice is given below:

- V_i - settling velocity of alga i (9 parameters),
- V_{POM} - settling velocity of particulate dead organic matter (1),
- $S_{N,i}$ - N:C ratio for alga i (12),
- $S_{P,i}$ - P:C ratio for alga i (12),

- S_{chlfi} - chlorophyll-a:C ratio for alga i (6),
- $S_{Si,i}$ - Si:C ratio for alga i (3),
- $P_{i,0}$ - maximal growth rate of algae type i at 0 °C (12),
- E_{Algi} - specific extinction of algae type i (12),
- $E_{Hum,0}$ - extinction due to humic substances in pure fresh water (1),
- K_b - background extinction (1),
- E_{SPM} - specific extinction of inorganic suspended matter (1),
- E_{POM} - specific extinction of particulate dead organic matter (1).

Chlorophyll is a light-sensitive pigment that is found in plants, algae and some bacteria. It can absorb light quanta therefore it is vital for photosynthesis, the process of converting light energy into chemical potential energy, followed by the fixation of inorganic carbon into sugars. It is known that the chlorophyll-a concentration is computed according to the following formula:

$$(20) \quad chlfa = \sum_{i=1}^n (s_{chlfi} * Alg_i)$$

where s_{chlfi} is a stoichiometry of chlorophyll-a in alga i and Alg_i is the computed biomass of alga i for $i=1,2,\dots,12$ corresponding to three different E,N,P types among four taxonomic groups: Diatoms, Flagellates, Dinoflagellates and Phaeocystis (description is given in Section 3). Different dry weight to chlorophyll-a ratios are used to describe the phenomenon that various types of phytoplankton have different preferred light intensities and abilities to adapt to new light or temperature conditions. However, the values of the chlorophyll-a ratio has been grouped together for different species of algae of the same type, excluding Diatoms, and this procedure resulted in a contribution of 6 input parameters. The justification of this approach is strongly based on laboratory experiments that have shown consistently different values for diatoms than for non-diatoms and for the latter there were no remarkable differences in chlorophyll-a content.

The change in algae biomasses can be described with means of the following processes:

$$(21) \quad \frac{\partial Alg_i}{\partial t} = gro_i - mrt_i - sed_i - grz_i$$

denoting respectively net phytoplankton growth, phytoplankton mortality, settling and grazing by filter-feeders. Moreover, algal sedimentation is defined as:

$$(22) \quad sed_i = \frac{v_i * Alg_i}{Z}$$

where Z stands for water depth, a model forcing. In case of Dino-Flagellates it is known that the sedimentation velocity is always equal to zero, hence it was not included in the parameter list. In the same manner one can describe detritus sedimentation:

$$(23) \quad sed_{POM} = \frac{v_{POM} * POM}{Z}$$

Linear programming is used as an optimization technique to determine the species composition that is best adapted to current environmental conditions. The suitability of a type is determined by the ratio of its requirement and its growth rate. Hence the maximum growth rates at 0 °C and other nutrient's stoichiometries namely N:C ratio, P:C ratio and Si:C ratio that is defined only for Diatoms, are used to determine Alg_i at each time step; their values are included in the constraints equations. More precisely the objective function reads:

$$(24) \quad \text{Max} \sum_{i=1}^n (pg_i * le_i - r_i) * Alg_i \quad \text{with respect to } Alg_i$$

satisfying the following constraints:

$$(25) \quad \sum_{i=1}^3 Alg_{i,new} \leq \sum_{i=1}^3 Alg_i * e^{(pg_i * le_i - r_i) * \Delta t} \quad (\text{growth constraint per species group})$$

$$(26) \quad \sum_{i=1}^3 Alg_{i,new} \geq \sum_{i=1}^3 Alg_i * e^{-m_i * \Delta t} \quad (\text{mortality constraint per species group})$$

$$(27) \quad \sum_{i=1}^n s_{N,i} * Alg_{i,new} \leq \sum_{i=1}^n s_{N,i} * Alg_i + NO_3 + NH_4 \quad (\text{nutrient constraint – nitrogen})$$

$$(28) \quad \sum_{i=1}^n s_{P,i} * Alg_{i,new} \leq \sum_{i=1}^n s_{P,i} * Alg_i + PO_4 \quad (\text{nutrient constraint – phosphorus})$$

$$(29) \quad \sum_{i=1}^n s_{Si,i} * Alg_{i,new} \leq \sum_{i=1}^n s_{Si,i} * Alg_i + Si \quad (\text{nutrient constraint – silicate})$$

$$(30) \quad k_{\min,i} \leq k_d \leq k_{\max,i} \quad (\text{light constraint})$$

where:

pg_i - maximal gross growth rate algae type i i.e. $pg_i = p_i + r_i$ where maximal net growth rate for algae type i , $p_i = p_{i,0} * (Temp - kt_{p,i})$,

le_i - growth efficiency of algae type i , tabulated function of light,

r_i - maintenance respiration rate for algae type I ,

m_i - mortality rate for algae type I ,

$k_{maz,i}$ - maximum extinction where the net growth of algae type i is positive; above this level self shading limits growth,

$k_{min,i}$ - minimum extinction where the net growth of algae type i is positive; below this level photo-inhibition limits growth,

k_d - total extinction coefficient.

As the sunlight penetrates the air-sea interface it is absorbed and scattered by the water molecules and water constituents, in particular phytoplankton whose cells contain chlorophyll-a. Extinction of light by substances in the water is modelled as an exponential decrease of light intensity with depth according to the Lambert-Beer formula (2). The total extinction coefficient in the last constraint is estimated as the sum of extinction by inorganic matter, particulate organic matter, phytoplankton, dissolved humid substances (approximated by salinity) and background extinction. Each of the substances is characterized by a specific extinction coefficient namely E_{SPM} , E_{POM} , E_{Algi} , $E_{Hum,0}$ and K_b . That justifies attaching great value to the parameters mentioned above. The uncertainty of the parameters can be found in Appendix. They were all assumed to be uniformly distributed over the certain ranges.

One model run lasts approximately 40 minutes and it is relatively short comparing to the 3D version. Therefore, to generate a Monte Carlo sample of one thousand 27 days of continuous simulation were required. For the method of Morris with 10 trajectories the total simulation time was a bit shorter namely 20 days. Additionally these two methods were applied simultaneously in order to save the waiting time. The output was produced every 7 days.

In order to analyze the obtained results it is essential to specify the output of interest, in other words, not the output as such but rather the question that the analysis has been called to answer. As it was mentioned in Section 2 the objective of this project is to investigate the chlorophyll-a concentration. However, there are many questions that one can ask: should it be done over the whole simulation time, particularly over the year, or only at the end, which locations should be considered, should one average over the time and/or locations? For future use, it was decided to deliberate the maximal concentration over the year, corresponding to algal bloom, together with the annual mean of concentration. With the intention to encompass specific characteristics of different areas, like the shore distance, river discharges, concentration of suspended matter, the total number of 49 monitoring stations was chosen and they are listed in the following table:

No	STATION	No	STATION	No	STATION
1	Walcheren02	18	Terschelling235	35	Terheijde4
2	Walcheren20	19	Rottumerplaat003	36	Terheijde30
3	Walcheren70	20	Rottumerplaat050	37	Terheijde70
4	Schouwen10	21	Rottumerplaat070	38	Appelzak1
5	Goeree06	22	Marsdiepnoord	39	Appelzak30
6	Noordwijk02	23	DooveBalgooost	40	Appelzak50
7	Noordwijk10	24	Vliestroom	41	Callantsoog1
8	Noordwijk20	25	DooveBalgwest	42	Callantsoog4
9	Noordwijk30	26	Blauwsot	43	Callantsoog20
10	Noordwijk50	27	Huibertgat	44	Callantsoog50
11	Noordwijk70	28	Vlissgbiss	45	Calandsoog 100
12	Terschelling004	29	Wissenkerke	46	Dantziggat
13	Terschelling010	30	Egmond1	47	Straat Dover
14	Terschelling050	31	Egmond10	48	Belgie-Engeland
15	Terschelling100	32	Egmond70	49	HANSWGL
16	Terschelling135	33	Harlingen		
17	Terschelling175	34	Terheijde1		

Table 5.1 Selected stations

The combination of the considered locations is representative for a wide range of conditions in the Dutch part of the North Sea. The following figure presents the locations of a few selected stations:

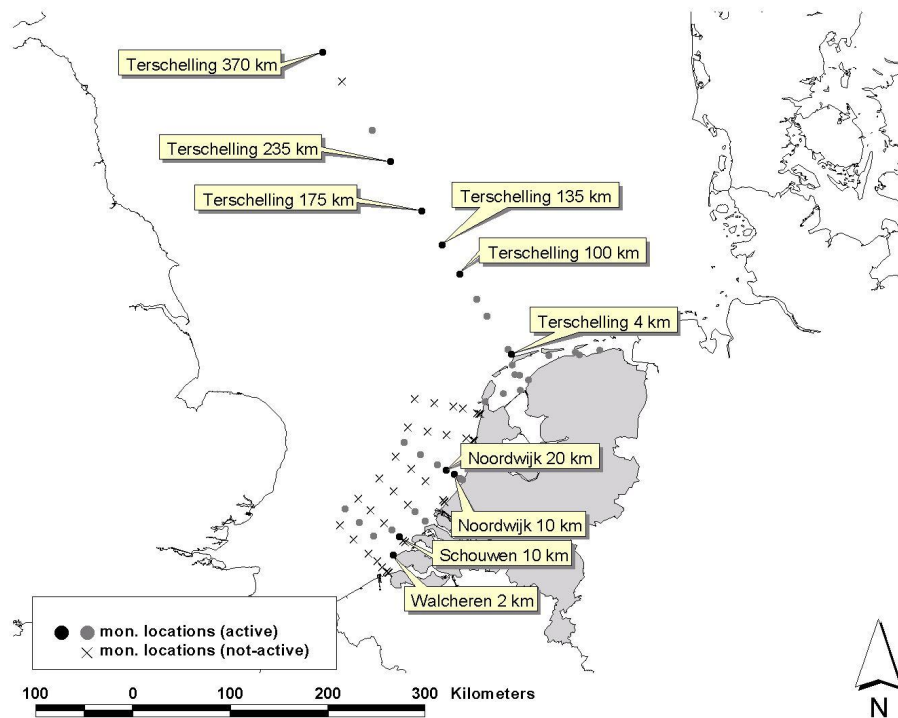


Figure 5.1 Locations of a few selected stations

It should be pointed out that uncertainty in the input was determined in an experimental way, hence certain ranges of parameters (all assumed to be uniform) might have been overestimated. Due to the fact that some parameter settings produced unrealistic results, for instance the bloom occurred in the winter, the blooming time was not taken into consideration, even if it seems to be essential for the water quality analysis.

It is advantageous to start with examining the result of a screening technique, in particular the Morris method. As it was mentioned before, (μ, σ) plane will be used for the sake of analysis. The interpretation of the meaning of μ (or μ^*) is straightforward since it corresponds to the main effect of a given factor. Hence the higher value μ , the stronger the influence on the output is. With respect to σ the interpretation is also intuitive: its high value indicates that the elementary effect of a given factor is strongly affected by the choice of other factor's values. In other words the elementary effects relative to this factor are significantly different for different input settings. Moreover, the potential non-linearities may cause high values of σ . On the other hand, a low value implies that the effect is almost independent of the values taken by the other factors and that the output is almost linear with respect to this factor.

Below the outcome of the first approach is presented. Parameters were denoted by numbers [App] and these numbers, put in parenthesis, will be further used in this report.

4.1 The Morris method - maximal chlorophyll-a concentration.

Let us start the analysis with looking at the results of the Morris method for station Schouwen10, situated 10 km offshore in the Delta region of the Netherlands:

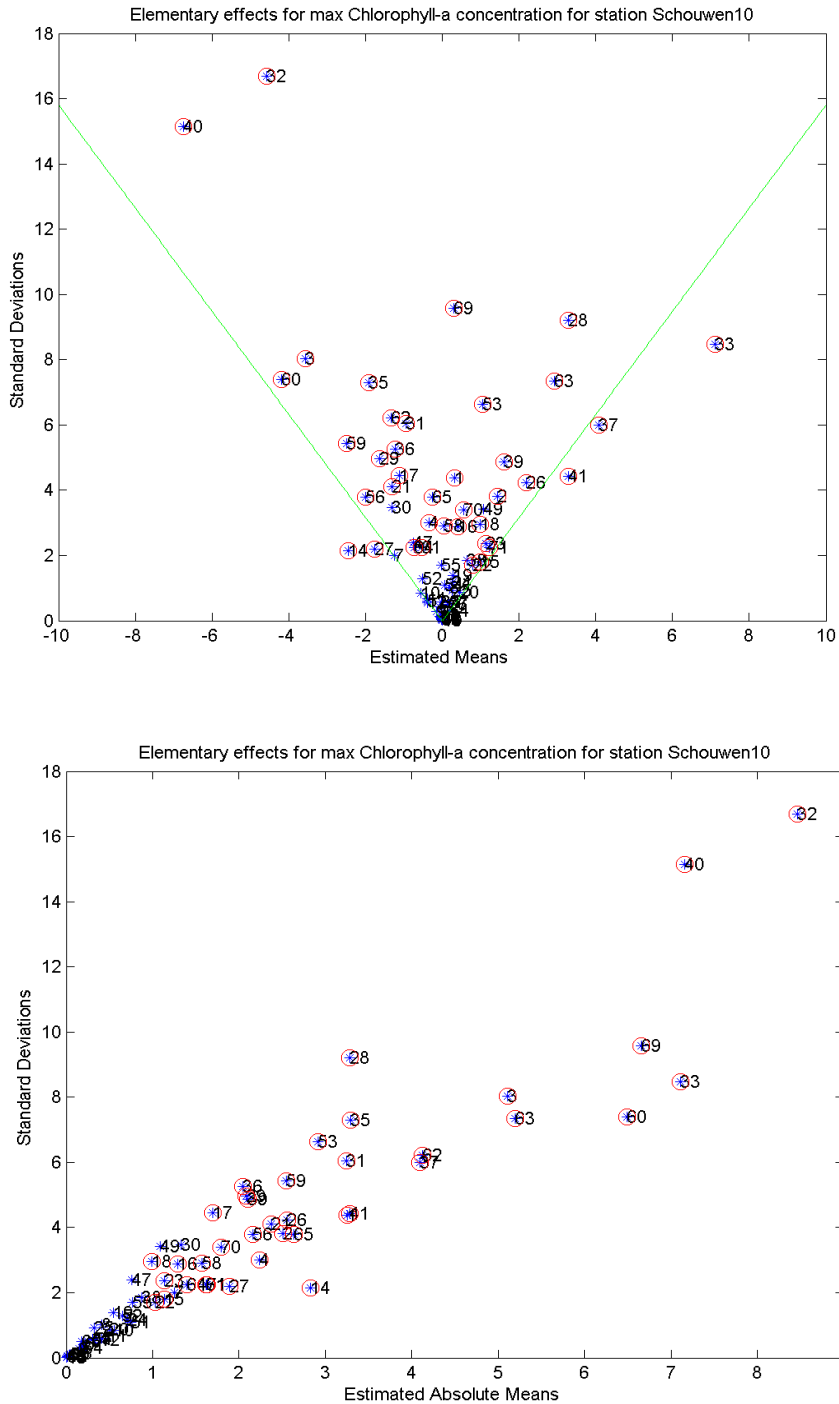


Figure 5.2 Top plot presents the means of the elementary effects plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Schouwen10. The bottom plot was made for the absolute values of the means.

Parameters that are well separated from the others for more than 10 stations are marked by red circles. Among them the fairly high expectations of elementary effects were obtained for the joint chlorophyll-a to carbon ratios denoted by 33, 37 and 41 and corresponding respectively to algae type E, algae type N and algae type P, each with exclusion of Diatoms. It is fully justifiable by the formula (20), where the chlorophyll-a concentration is computed as the sum of products of ratios and the current algae biomasses. Moreover, their effect on the chlorophyll-a concentration is non-linear since the latter are found by the optimization technique where other input parameters were used in the definition of constraints. In Figure 5.2 one can notice high values of the standard deviations not only for the chlorophyll-a to carbon ratios, but also for other factors, that appeared due to existing interactions.

The top graph clearly shows that parameters 41, 33 and 37 are lying outside the wedge. Chlorophyll-a to carbon ratios for Diatoms (18, 23 and 28) are still important but certainly less than their counterparts. This can be explained by the fact that the change of one of the joint parameters follows actually the simultaneous change of three parameters for three different species of algae of the same type. However, it is not always the case, especially if the location of interest is dominated mostly by Diatoms, for instance Noordwijk30. Then the influence of 18, 23 and 28 increases remarkably, that is illustrated in the following figure:

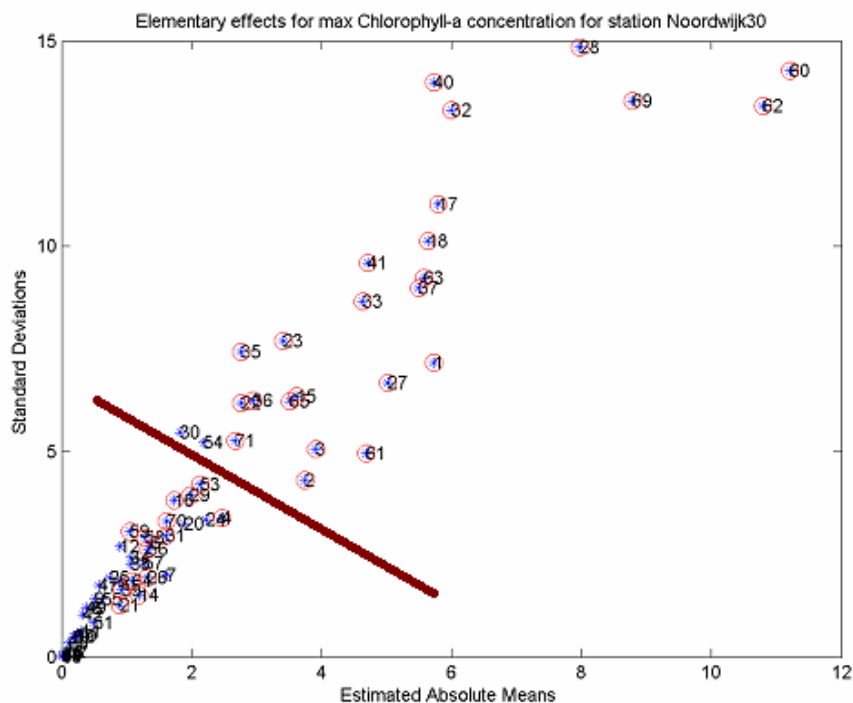


Figure 5.3 The absolute values of means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Noordwijk30.

Returning to Figure 5.2, one can notice that factors 32 and 40 denoting respectively phosphorus to carbon ratios for Flagellates type E and type P play a significant role, however they are either involved in interactions with other parameters or have highly non-linear effects on the output (value of σ is high). Moreover, these effects are of different signs, depending on the point of the space at which the effect is computed. It should be explained in the following way: the increase of one of the mentioned ratios, corresponding to requirements for algae growth, may lead to the smaller total algae biomass that will finally result in a smaller bloom and with a different parameter setting it can cause the increase of algae bloom. Our analysis shows how beneficial the use of μ^* can be. Interestingly, values of these two factors are currently being recalibrated because of extremely high sensitivity of Flagellates to phosphorus. For Terschelling transect, and for a few other stations, they were found less important probably due to the low value of phosphorus concentration and high value of nitrogen concentration that are typical of stations situated along this transect. In the same manner the impact of maximum growth rates denoted by numbers from 60 to 71 is twofold, a rise of one of them can lead to a higher peak in the chlorophyll-a concentration if the environment conditions are favourable but on the other hand it also affects the process of algae adaptation hence a new algae composition and that strongly depending on the values of other parameters can result in the reduction of size of the bloom. This fact is clearly pictured in the Figure 5.2 demonstrating high values of the standard deviations and significant change in the expectations between the top and bottom plot. Additionally, growth rates of all types of Dino-Flagellates (66, 67 and 68) were shown to be negligible, as well as the other factors corresponding to this species and that can be justified by a low concentration of this group of algae found in the North Sea. An exception to this observation is Terschelling transect where in July exactly this species becomes dominant, therefore the variation of the values of the parameters related to Dino-Flagellates could cause their sooner appearance and possibility of affecting the size of bloom. This appears at Terschelling235 as shown in the following figure:

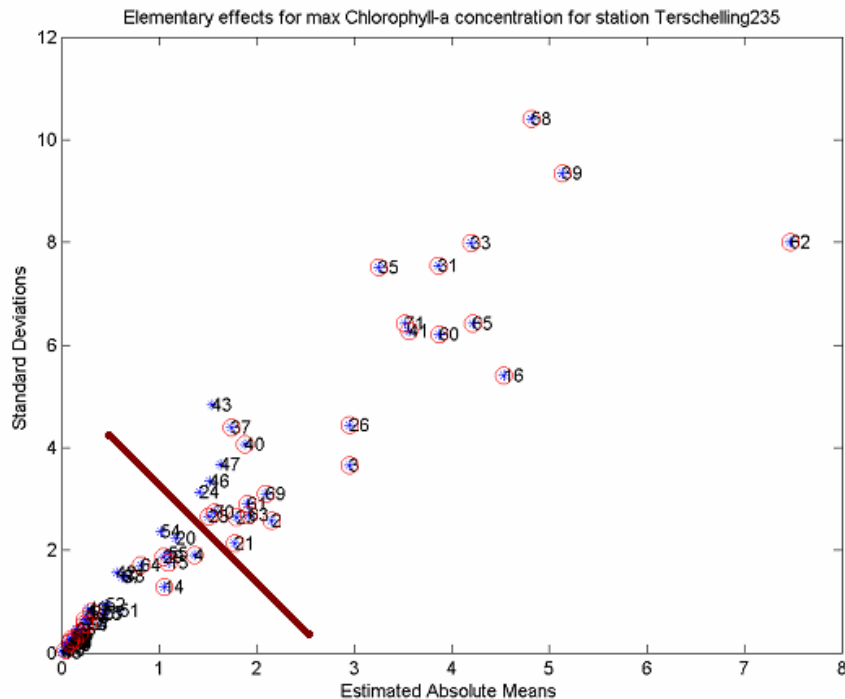


Figure 5.4 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Terschelling235.

P:C ratio of Dino-Flagellates type N (47) appeared among the influential factors more often, but always on a small scale (the values of mean and standard deviation were relatively small comparing to other parameters).

As it was mentioned before, primary production is strongly affected by light availability. Therefore, it is not surprising that parameters related to light extinction turned out to affect the size of bloom, but again with effects of different signs. The more light is available the more efficient production is. However, it can be used by different types of algae and it can become extinct due to different substances in the water, that further can cause either reduction or increase of the maximal chlorophyll-a concentration. Specific extinction due to humic substances in pure fresh water (1) and specific extinction of inorganic suspended matter (3), produced quite high expectations of the elementary effects that can be easily noticed on the bottom of Figure 5.2. A bit less important seems to be background extinction (2) and extinction of dead particulate matter (4). Although, for different locations, for example at Terheijde70 characterized by a small amount of suspended matter and a relatively long distance from the shore, it was observed that (2) and (4) are even more significant. That is illustrated in the figure below:

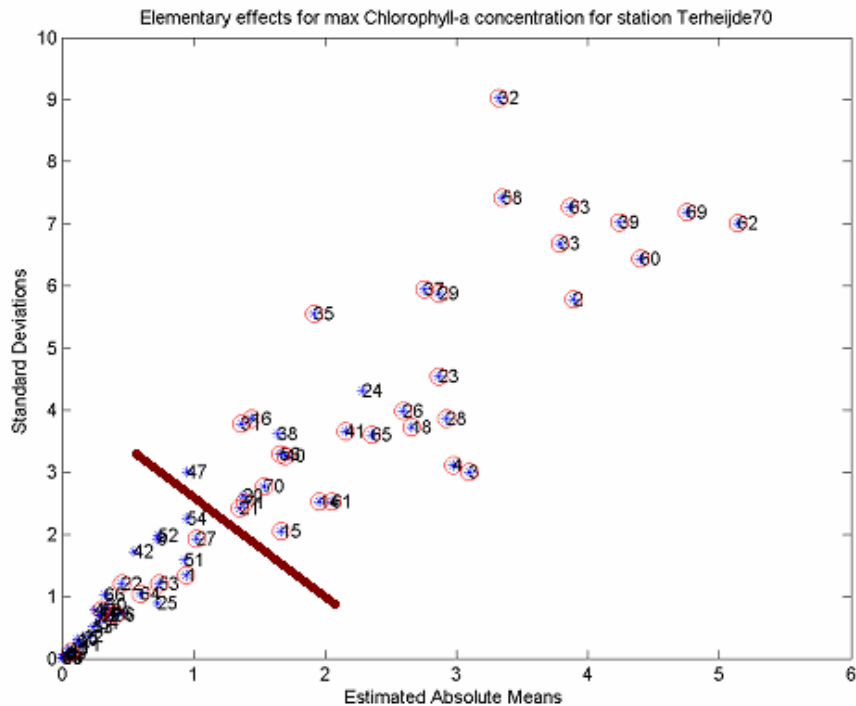


Figure 5.5 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Terheijde70.

Intuitively, the growth of importance of 1 is mostly observed in the areas close to the shore and/or influenced by river plums, in particular at Goerre6 affected by Haringvliet, Harlingen affected by Lake IJssel and both Terheijde and Noordwijk affected by Rhine. It can be concluded that light plays a key role for the size of bloom.

When analyzing parameters related to algae requirements, in particular P:C, N:C and Si:C ratios it is hard to make a general statement about the impact of the respective uncertainty on the chlorophyll-a concentration. At different places on the North Sea the leadership is taken by different sets of factors depending on the local characteristics and the set that is considered as influential at one station might not be deciding at another. For example, at Noordwijk70, a place of high concentration of Diatoms, Si:C ratios for types N and P (24 and 29) as well as specific extinction of all the types (15, 20 and 25) were recognized as not negligible. The same was observed for only a few other stations, in particular within this transect:

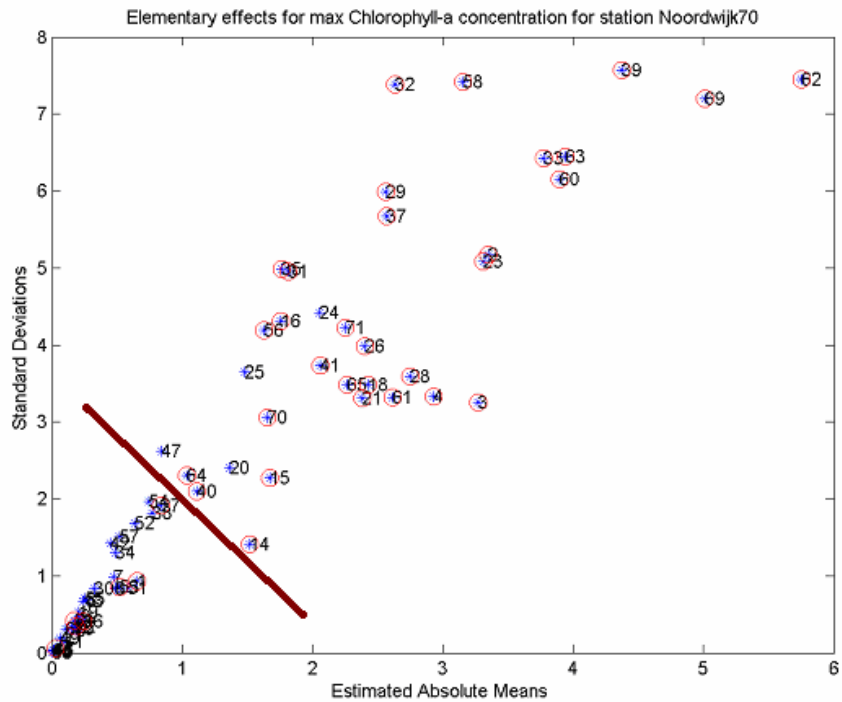


Figure 5.6 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Noordwijk70.

It is worth noting that factors 69, 58, 56 and 70 corresponding to *Phaeocystis* characteristics were also pointed out as of high importance, since quite high concentrations of this species are specific for the considered area.

In the same manner, at DooveBalgooost uncommonly indicated parameters are settling velocity of Diatoms type N (6), P:C ratio of Diatoms N (22) and P:C ratio of Diatoms type E (17). The latter seems to be influential for approximately half of the considered locations.

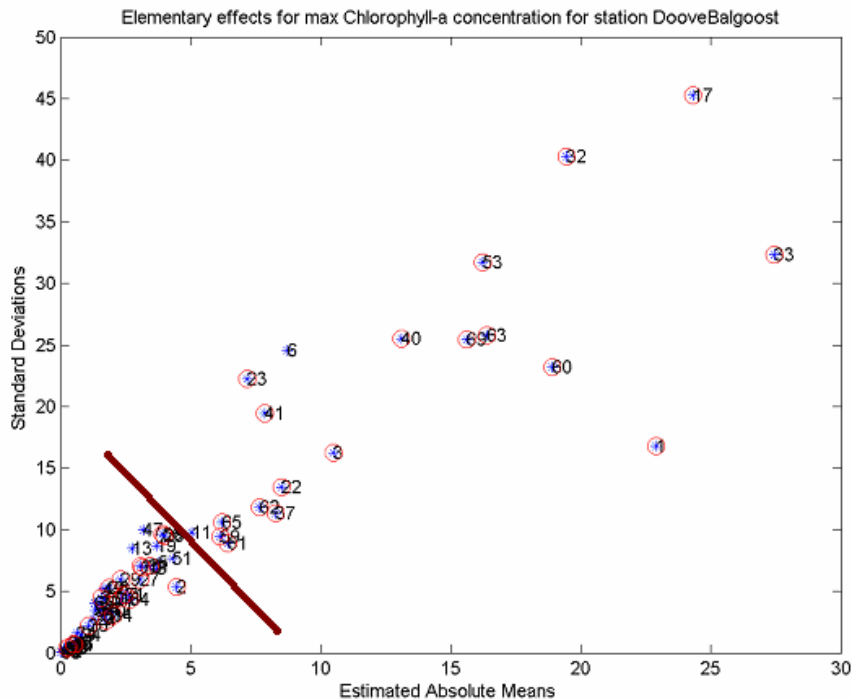


Figure 5.7 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station DooveBalgooost.

It should be stressed here that this station is situated in the Wadden Sea and as it was mentioned before it is a place where model performance is rather poor because of the steep gradients of sediments and nutrients.

Summarizing, the analysis of the entire number of plots showed that P:C and N:C ratios of Flagellates and a bit less but still of Diatoms should be regarded as important. For Phaeocystis it turned out to be right only for P:C ratios of all types and N:C ratio in case of type P. It is worth noting that a great influence of N:C ratios was observed at Terschelling175 (characterized by high nitrogen contents and stratification), in particular for Diatoms type E, Diatoms type N, all types of Flagellates and Phaeocystis type N (16, 21, 31, 35, 39 and 58). It is shown in the following figure:

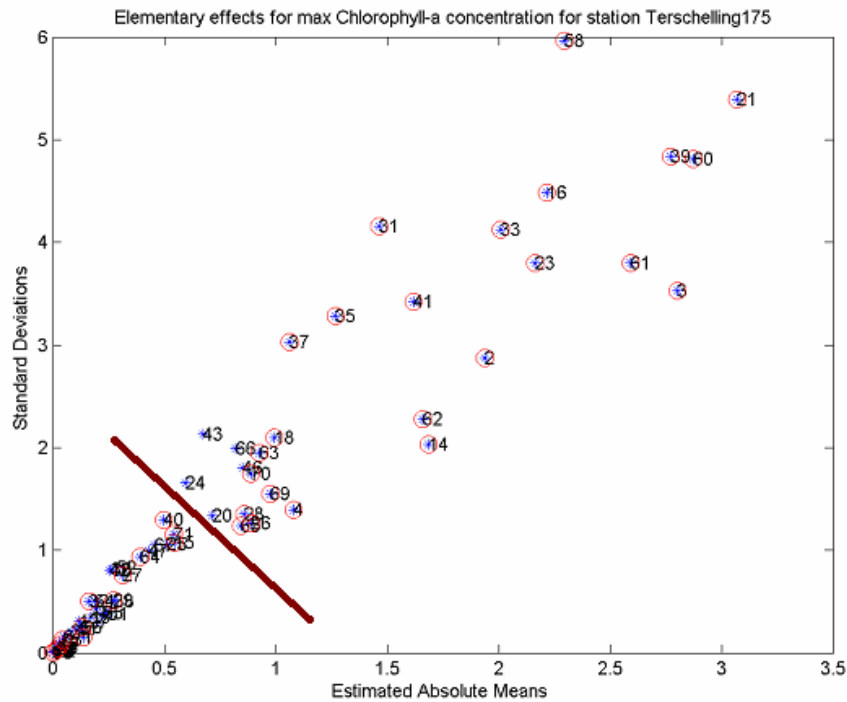


Figure 5.8 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Terschelling175.

One can also notice that the importance of factors 32 and 40 decreased, which is in agreement with the previous statement about phosphorus contents.

Other parameters, for example specific extinction of all types of Flagellates and Phaeocystis, seem not to play a substantial role in the change of the chlorophyll-a concentration, elementary effects computed for these parameters have relatively small absolute means for most of the concerned stations. High values of σ for all the parameters demonstrate that interactions play an important role in the model. It can also be concluded that none of the factors has a purely linear effect, since all the points lie around the diagonal. However, it is worth noting that sedimentation velocity for detritus (14) was found as a factor with the most linear effect on the output and quite high value of the mean at the same time. Sedimentation velocities for different types of algae are not very important with respect to the others, they do not influence the magnitude of fluctuations of algae blooms.

We see that results produced for different locations are very similar. Although, due to various characteristics recognized for these locations, the order of importance is not always the same and some parameters at some points influence the output stronger than at the others.

The above discussion can be summarized in the following barplot, representing the number of stations, where a given parameter was indicated as important by the Morris method. In other words, the point corresponding to this parameter was well separated from the others in the (μ^*, σ) plane, i.e. it was lying behind the line that was marked in Figures 5.3-5.8.

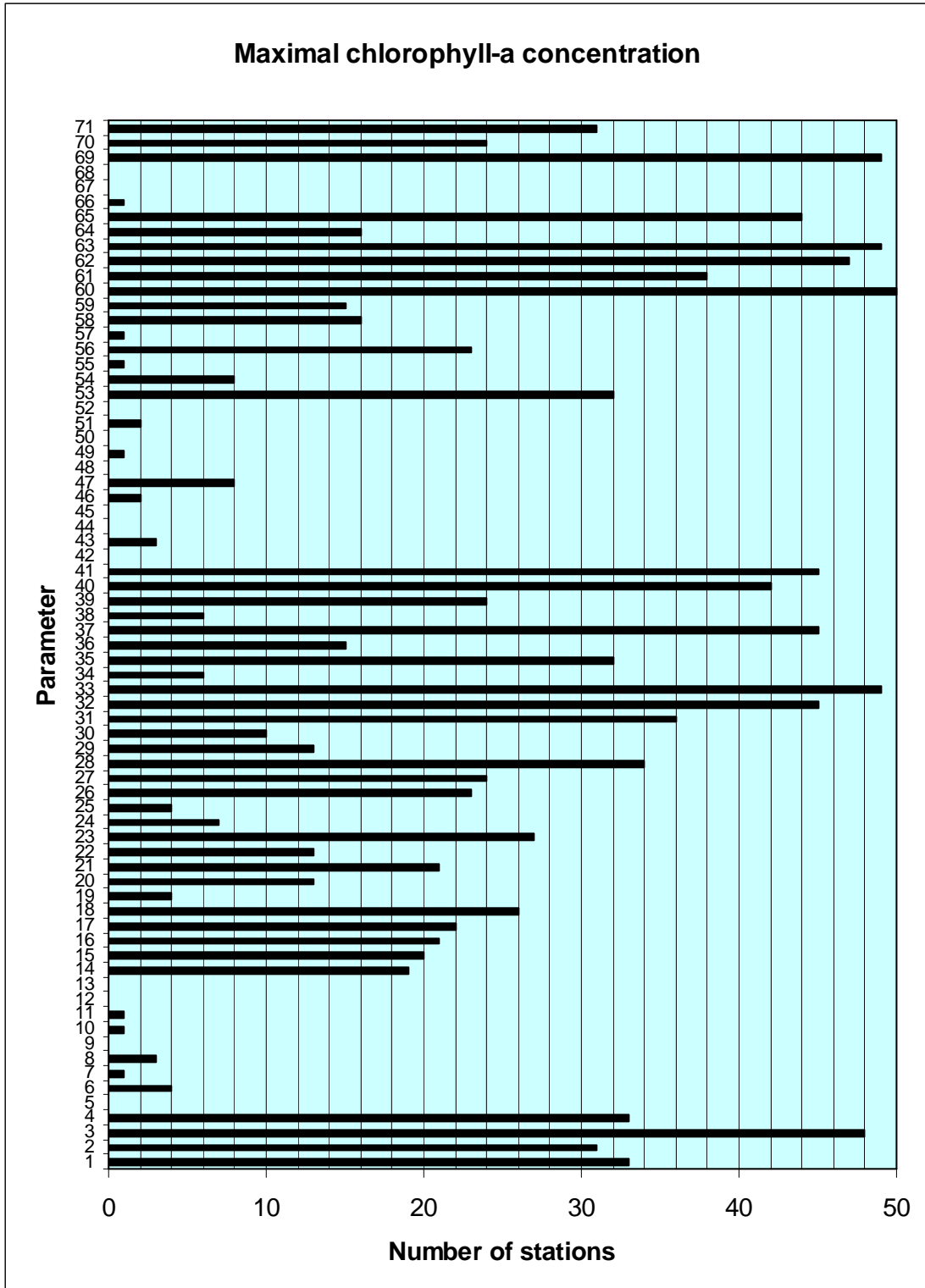


Figure 5.9 The number of stations where a given parameter was indicated as important.

We see that parameters 3, 33, 60, 62,63 and 69 are overall the most influential factors, their importance was pointed out at more than 45 stations.

The Euclidean distance is sometimes used in assessing sensitivity effects. To get better insight into the order of importance of the analyzed parameters at each location j the following measure was computed for each of them:

$$(31) \quad \varepsilon_{i,j} = \sqrt{\mu_{i,j}^2 + \sigma_{i,j}^2} \quad \text{for } i=1,\dots,71 \quad j=1,\dots,49$$

and then the ranks were assigned to the parameters on the basis of the Euclidean distances. More precisely, a rank from one (highest) to 71 (lowest) for each parameter. Next they were summed up over all the stations to make a compaction and comparison of the data possible. The results are presented in the following barplot:

Sums of ranks for maximal concentration

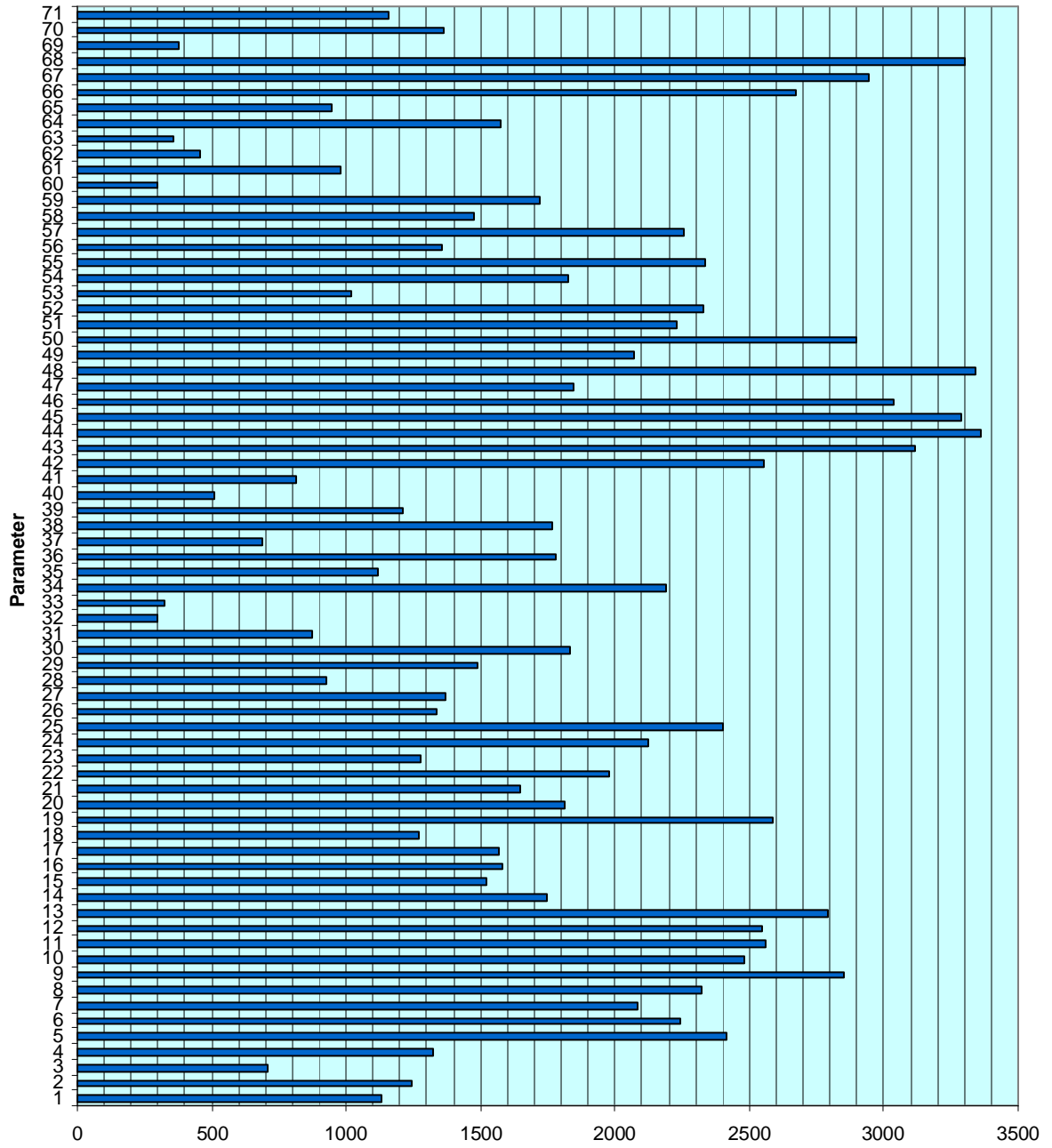


Figure 5.10 Sums of ranks of parameters 1,...,71 over all the stations computed on the basis of the Euclidean distance. Low value indicates high importance of a given parameter.

Among them, 20 that obtained the highest ranks are listed below:

	Number	Parameter:	Rank:
1	60	maximum growth rate at 0 °C for Diatoms type E	299
2	32	P:C ratio for algae Flagalletas type E	300
3	33	chlorophyll-a:C ratio for algae type E	326
4	63	maximum growth rate at 0 °C for Flagalletas type E	356
5	69	maximum growth rate at 0 °C for Phaeocystis type E	377
6	62	maximum growth rate at 0 °C for Diatoms type P	455
7	40	P:C ratio for algae Flagalletas type P	512
8	37	chlorophyll-a:C ratio for algae type N	685
9	3	specific extinction of inorganic suspended matter	709
10	41	chlorophyll-a:C ratio for algae type P	816
11	31	N:C ratio for algae Flagalletas type E	875
12	28	chlorophyll-a:C ratio for algae Diatoms type P	927
13	65	maximum growth rate at 0 °C for Flagalletas type P	944
14	61	maximum growth rate at 0 °C for Diatoms type N	979
15	53	P:C ratio for algae Phaeocystis type E	1020
16	35	N:C ratio for algae Flagalletas type N	1119
17	1	extinction due to humic substances	1132
18	71	maximum growth rate at 0 °C for Phaeocystis type P	1161
19	39	N:C ratio for algae Flagalletas type P	1213
20	2	background extinction	1247

Table 5.2 20 parameters that obtained the highest ranks

4.2 The Morris method - annual average chlorophyll-a concentration.

Now let us turn to the analysis of change in the annual average of chlorophyll-a concentration. Obviously, the parameters that caused a noticeable change in the size of bloom should also influence on the annual average. However, one can observe small discrepancies in the way they affect a new output function. Once more, chlorophyll-a to carbon ratios of algae of different types (33, 37 and 41) were indicated as strongly influential, but this time 33 was dominant in that sense that it was lying evidently farther for almost all the stations. Single Chla:C ratios of Diatoms, in particular of type P, could be still consider as important but their influence on the annual average has never been high comparing to the other pointed parameters. In order to give insight into the new elementary effect's behaviour and compare it with the previous case the following figure, made for Schouwen10, is presented:

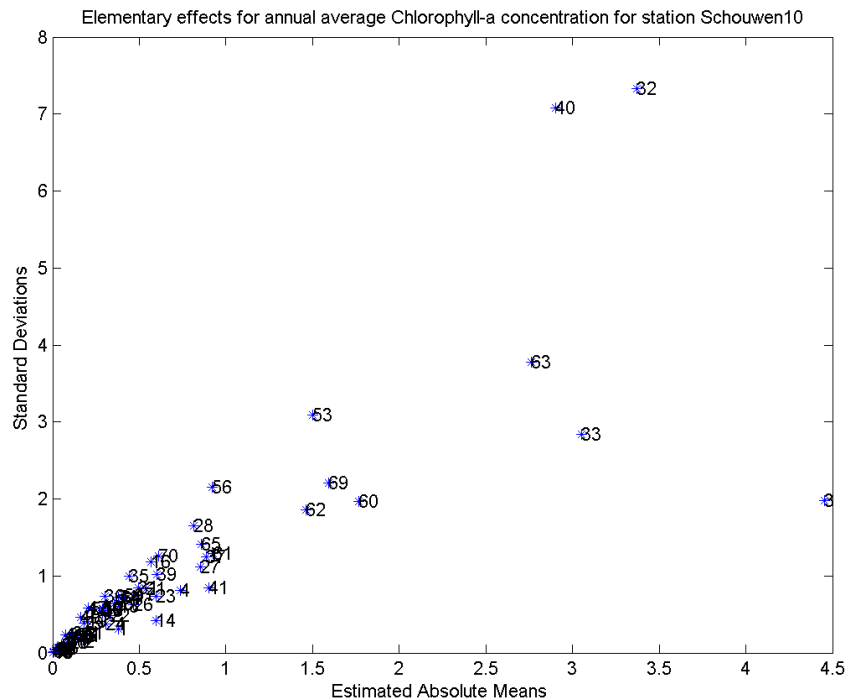


Figure 5.11 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Schouwen10.

It was noticed that again 32 and 40 caused big standard deviations of the elementary effects computed for these parameters for most of the stations, excluding again Terschelling transect, but within these two P:C ratio of Flagellates type E obtained a higher value, which can suggest a greater role of light in the long-run influence on the chlorophyll-a concentration. Furthermore, the growth rates of all types of algae, without the ones specified for Dino-Flagellates, are of crucial importance. It is worth noting that there are many similarities in Figures 5.2 and 5.11, however a few parameters were excluded from consideration for this particular station, for instance 31, 35 or 1 and others appeared to affect the new output function. The presence of P:C ratios of Phaeocystis type E and type N (53 and 56) among the most influential factors, well separated from the rest of parameters, was confirmed also by other plots showing the long-run impact of this species on the chlorophyll-a concentration. All the parameters linked to light appeared to be relatively important. The highest mean of the elementary effects can be observed for specific extinction of inorganic suspended matter (3) and specific extinction of humic substances (1), that is typical of stations close to fresh water discharges, but the first one to a greater degree. The latter is still dominant at such stations as Harlingen, Egmond10, Noordwijk10 or Terheijde4 but for instance at Goeree6 its influence

was significantly reduced that can be observed in the Figure 5.12 and at Walcharen2 its effect was even totally vanished that is showed in the Figure 5.13:

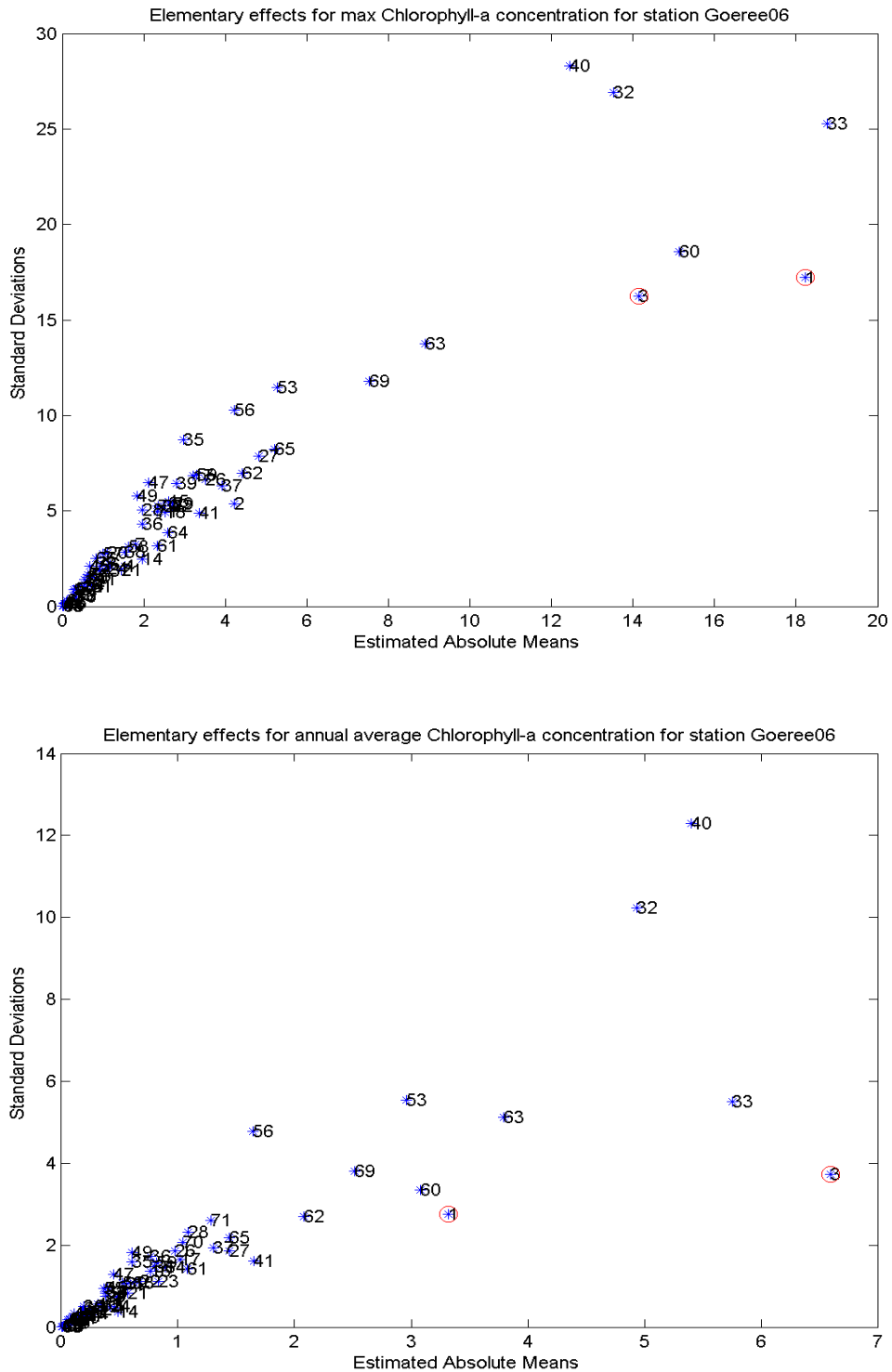


Figure 5.12 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Goeree06. On top for the maximal, on the bottom for the annual average chlorophyll-a concentration.

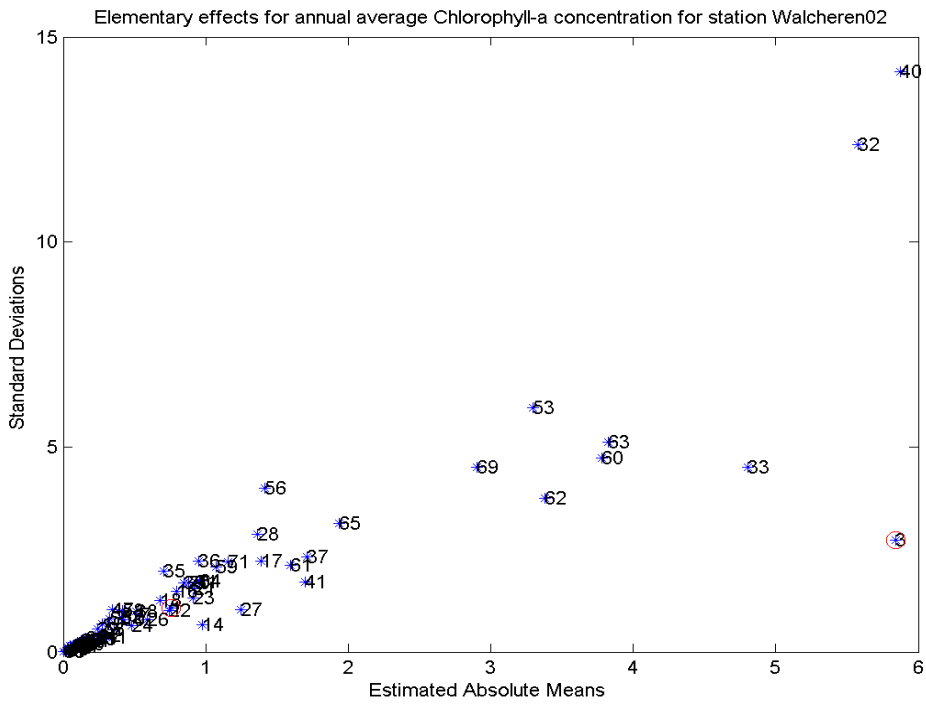
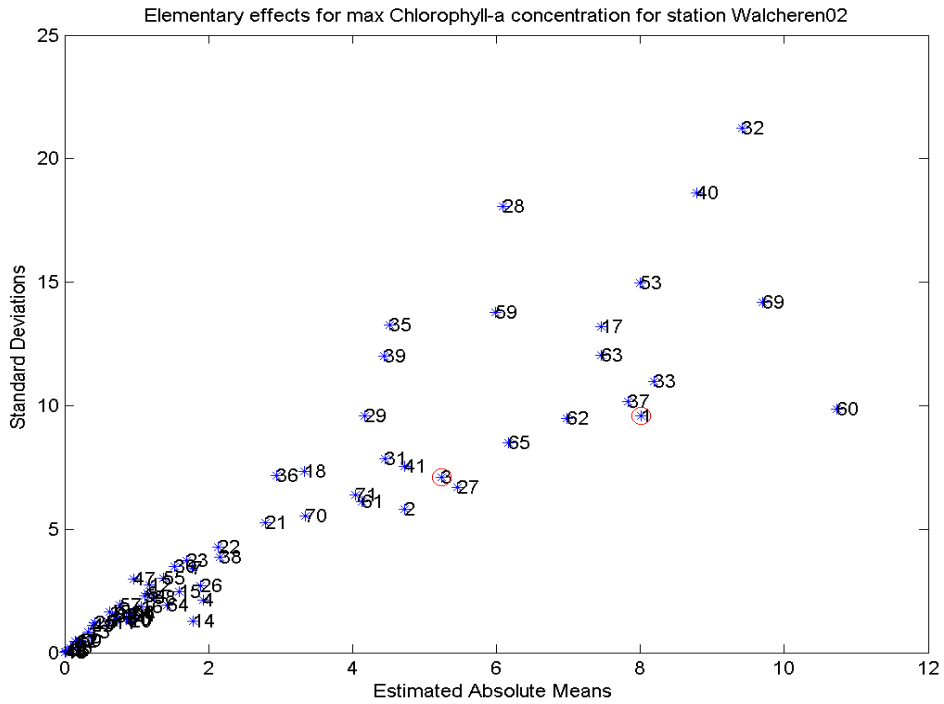


Figure 5.13 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Walcheren02. On top for the maximal, on the bottom for the annual average chlorophyll-a concentration.

The effect of change of background extinction (2) was found as not negligible at some locations but this factor does not affect the output as strongly as in the previous case, even at Terheijde70:

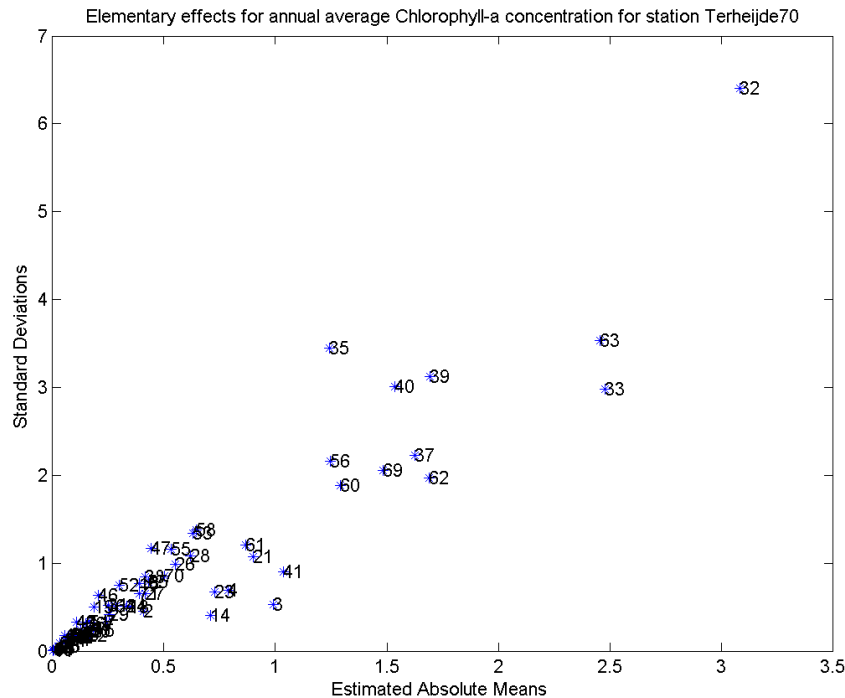


Figure 5.14 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Terheijde70.

What is more, single specific extinction of each type of alga turned out to be insignificant for all the stations.

Also this time, the strength of effect of change in P:C and N:C ratios was dependent on the location and its characteristics. It can be concluded that for Diatoms and Flagellates these parameters are quite important as well as for Phaeocystis but with exclusion of N:C ratio of type E (52) and P:C ratio of type P (59) and as it was mentioned before, with particular emphasis on importance of 53 and 56. Similarly, at Terschelling175 N:C ratios play a crucial role that is shown in the following figure:

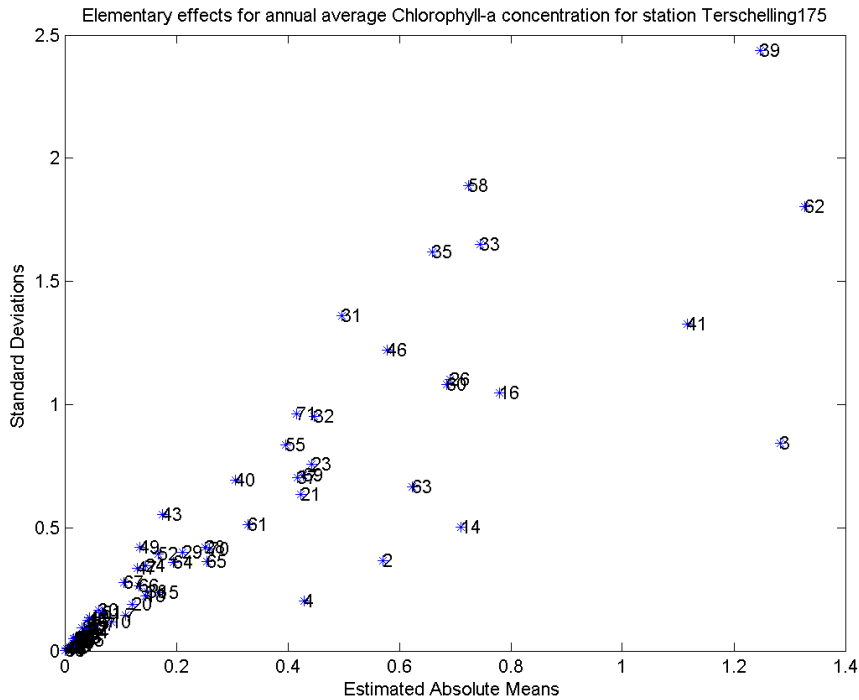


Figure 5.15 The absolute values of the means of the elementary effects are plotted against the standard deviations of the elementary effects computed for parameters 1,...,71 at station Terschelling175

It would not be surprising if the parameters related to algae that were not mostly responsible for an increase or decrease of the maximum concentration, or in other words, that did not dominate the species composition during the spring bloom turned out to be meaningful in case of the average concentration. However, once again Dino-Flagellates characteristics were found fairly insignificant for almost all the stations. Only N:C ratio type N (46) can be considered as one of factors leading to significant change in the annual average, but only at distant stations within Terschelling transect..

Eventually, other parameters like Si:C ratios of Diatoms or settling velocities, without the one specified for detritus, can be considered as negligible in affecting the annual average concentration of chlorophyll-a.

The summary of the above discussion is yet again presented in the barplot:

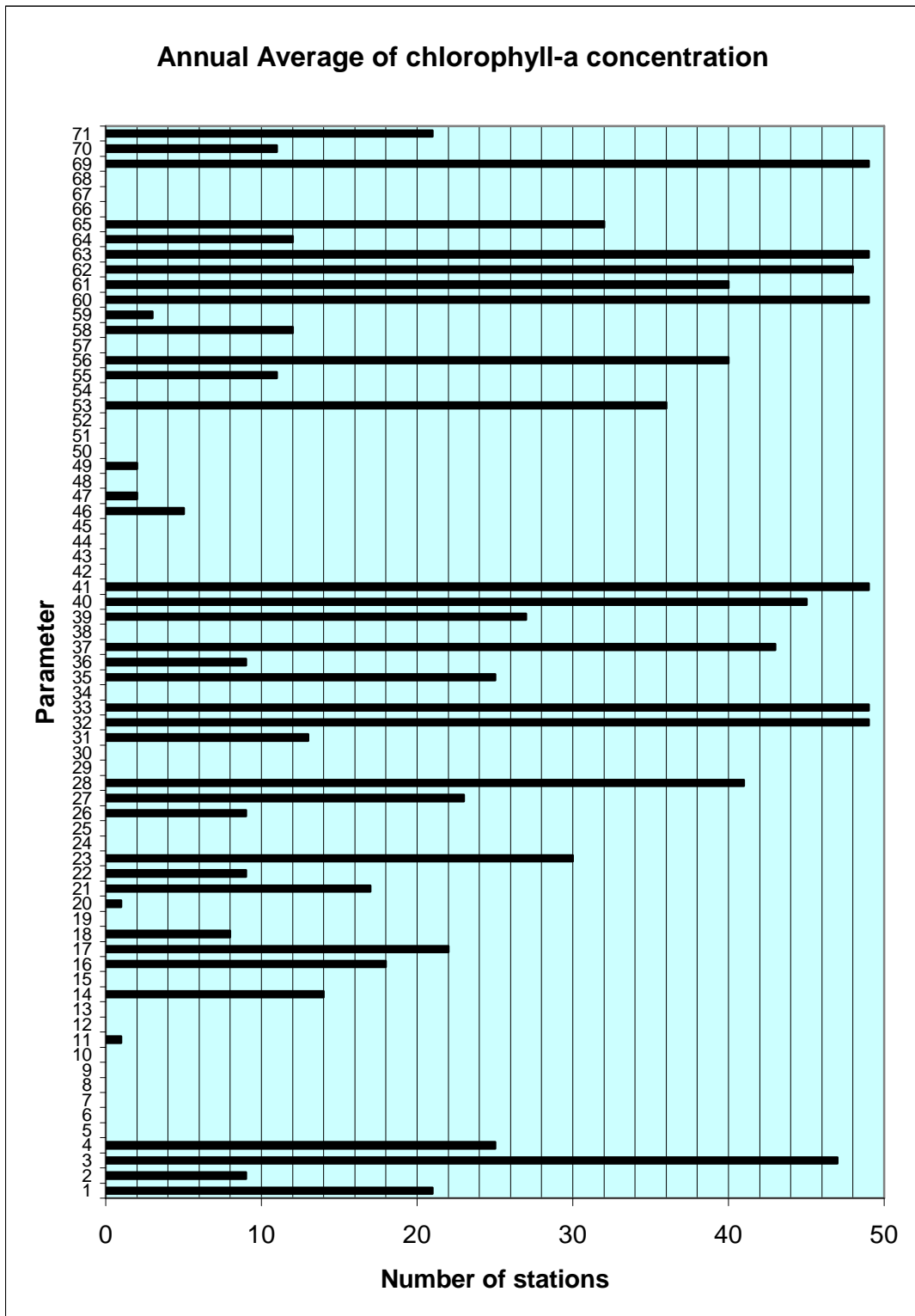


Figure 5.16 The number of stations where a given parameter was indicated as important.

One can have a look as well at the ranks of parameters computed on the basis of the Euclidean distance between the point (0,0) corresponding to no influence and the point (μ^*, σ) corresponding to the level of importance of a given parameter:

Sums of ranks for annual average

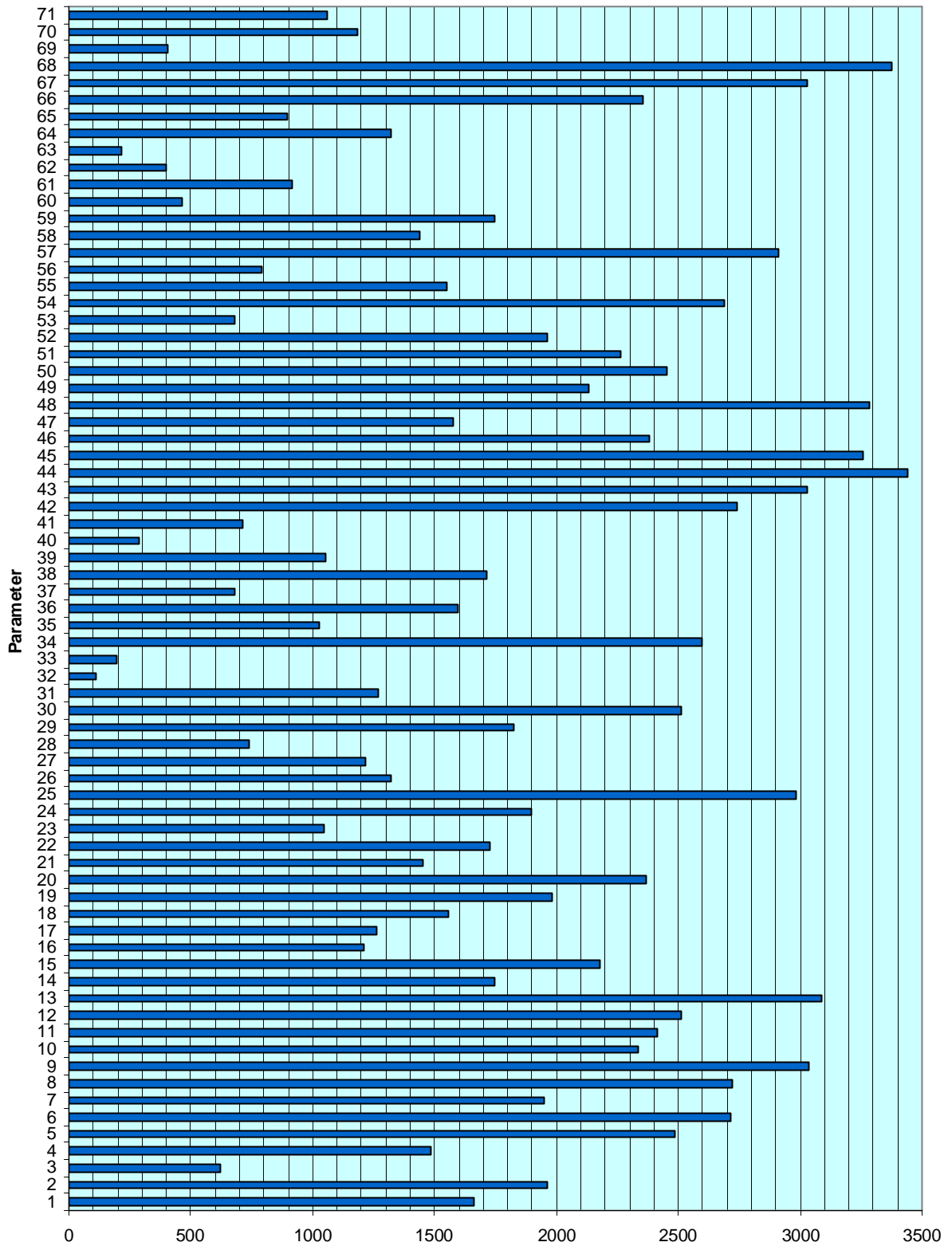


Figure 5.17 Sums of ranks of parameters 1,...,71 over all the stations computed on the basis of the Euclidean distance. Low values indicate high importance of a given parameter.

The following table shows a great number of similarities with the previous case. However, the minimal rank is noticeably smaller, that shows a lot of common patterns in the order of importance over the stations for the annual average concentration. Maximum growth rate of Diatoms type E (60) dropped from the first down to the 7th position, but P:C ratio of Phaeocystis type E (53) jumped up from 15th to 10th position. Moreover, new parameters showed up among the 20 most influential for, namely 56,23 and 70 and as it was mentioned before the role of light became less major causing exclusion of 1 and 2 from this group. N:C ratio of Flagalletas type E (31) lost its importance as well, but it is still among 25 the most influential.

	Number	Parameter	Rank
1	32	P:C ratio for algae Flagalletas type E	109
2	33	chlorophyll-a:C ratio for algae type E	194
3	63	maximum growth rate at 0 °C for Flagalletas type E	217
4	40	P:C ratio for algae Flagalletas type P	288
5	62	maximum growth rate at 0 °C for Diatoms type P	401
6	69	maximum growth rate at 0 °C for Phaeocystis type E	406
7	60	maximum growth rate at 0 °C for Diatoms type E	467
8	3	specific extinction of inorganic suspended matter	622
9	37	chlorophyll-a:C ratio for algae type N	679
10	53	P:C ratio for algae Phaeocystis type E	680
11	41	chlorophyll-a:C ratio for algae type P	711
12	28	chlorophyll-a:C ratio for algae Diatoms type P	738
13	56	P:C ratio for algae Phaeocystis type N	792
14	65	maximum growth rate at 0 °C for Flagalletas type P	896
15	61	maximum growth rate at 0 °C for Diatoms type N	916
16	35	N:C ratio for algae Flagalletas type N	1026
17	23	chlorophyll-a:C ratio for algae Diatoms type N	1045
18	39	N:C ratio for algae Flagalletas type P	1052
19	71	maximum growth rate at 0 °C for Phaeocystis type P	1061
20	70	maximum growth rate at 0 °C for Phaeocystis type N	1182

Table 5.3 20 parameters that obtained the highest ranks

4.3 Variance-based analysis - maximal chlorophyll-a concentration.

Another method that was used to detect sensitivities in the model was the variance-based analysis. It does so by quantifying the contribution that each input factor makes to the variance of the output of interest. The reason for applying this method follows from the fact that the Morris method does not explore the effect of change in the value of a given parameter within the range between the previously specified levels and tells the analyst only a little about the possible correlations between the factors. It does not investigate small changes in the

parameter value. Naturally, these two methods examine sensitivities in the model with means of different measures, therefore one should be aware of the fact that they can produce different results. The Morris method for each input factor picks r points in the input space, corresponding to the number of trajectories, and computes the incremental ratios at these points. Later, they are averaged to compute the main effect and additionally the standard deviation is calculated to check non-linear effect. The second method decomposes the variance of the output according to the input variables. Obviously, in case of linear uncorrelated models the results of these two methods are the same. As it was mentioned before, a special module within Unicorn, called Unisens, was used in order to compute the correlation ratios, corresponding to the main effects of the selected parameters, and several other measures of uncertainty contribution, in particular the product moment correlations, rank correlations and regression coefficients. For the estimation of conditional expectation in the formula (13) a polynomial fit procedure was used. The analysis of GEM is conducted on the basis of the correlation ratio of the output with a given factor. Other measures, like the product moment correlation or rank correlation between the input and a given factor, are less reliable due to their ability to check only monotone relationships.

In the first place, let us analyze the parameters' behaviour at Schouwen10. Below the variance decomposition of the maximal concentration of chlorophyll-a is presented:

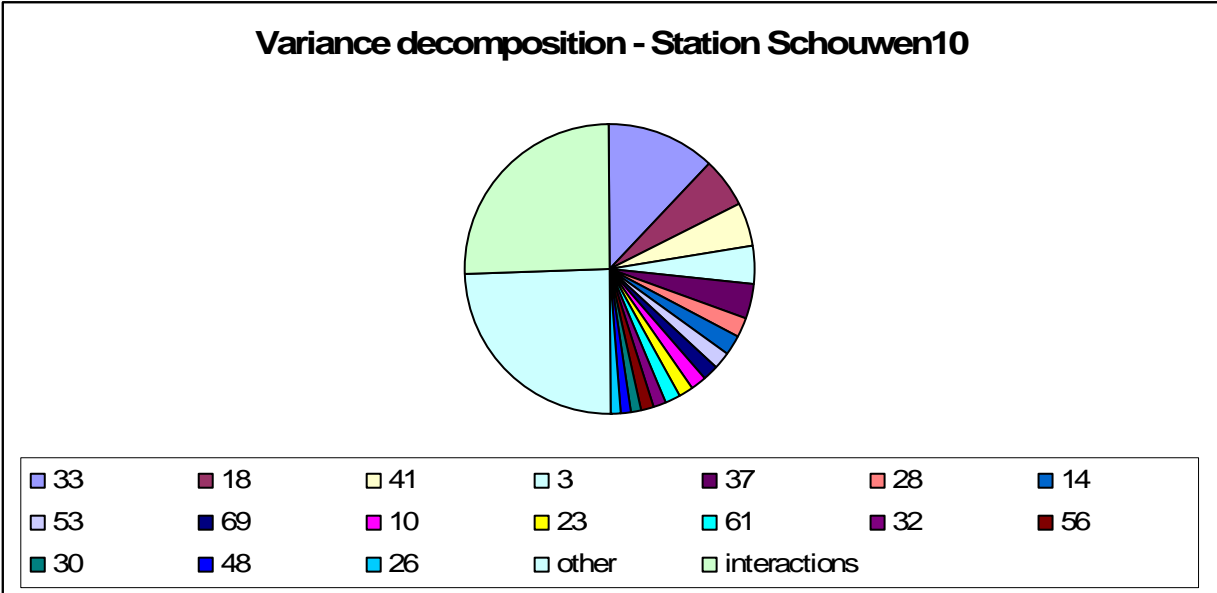


Figure 5.18 Variance decomposition of the output at station Schouwen10

Due to the fact that the contribution of interaction terms between the parameters was not taken into account it was not possible to recover 100 % of the variance of the output. The sum of parameters' contributions smaller than 1 % was denoted by 'other'. Once again, one can notice a great role of chlorophyll-a to carbon ratios of all types of algae. At this station parameter 33 is dominant and it covers approximately 12 % of the variance. Obviously, the order of importance of Chlfa:C ratios is different at different locations and depends on the type of algae that takes a leadership in the spring chlorophyll bloom. However, for almost all stations 18, 23, 28, 33, 37 and 41 appeared among the first fifteen the most influential factors.

Now let us turn to the light related parameters. Clearly, specific extinction of inorganic matter (3) has a major influence on the size of bloom. Moreover, this location is not affected by fresh water discharges, hence parameter 1 appeared among the group of potentially negligible factors. But again, at some stations it turned out to be of crucial importance, in particular Goeree6, Noordwijk, Rottumer, Harlingen and the stations situated around it, Terheijde and Vliessgbiss. The pie-plots of variance decompositions of the maximal chlorophyll-a concentration at the stations situated along Terschelling, Noordwijk, Terheijde and at Goeree06 are included in Appendix.

Similarly as the results produced by the Morris method parameters 2 and 4 were found less influential at this particular location. Generally, the importance of background extinction increased for example at Terheijde, Terschelling, a bit to the East Rottumer and Appelzak all situated farther from the shore, Noordwijk70 and situated next to it Egmond70 or Walcharen70 and it is more or less in agreement with the results of the first approach. Specific extinction of detritus (4) came out to influence the maximal concentration the most at Wissenkerke, but still the contribution to the output variance was low, namely 1.07 % and its effect was not detected by the Morris method.

The variance-based analysis emphasized the sensitivity of the maximal chlorophyll-a concentration to settling velocity of detritus (parameter 14). At Schouwen10 it is the 7th the most important parameter. At 40 other stations it was pointed out as member of the group of parameters with the highest contribution, but generally with a relatively smaller influence comparing to other members (the strongest at Walcheren20, around 2 %).

P:C ratio of Flagellates type E (32) seems to be less significant to the change in the size of bloom (1.5 %) and the correlation ratio of P:C ratio of Flagellates type P (40) turned out to be quite small, in particular less than 1%. Therefore, it was decided to take a closer look at the other mathematical measures for this parameter, but neither the value of product moment correlation nor rank correlation between the output and factor 40 was significantly high. The

regression coefficient of the output on 40 computed with the use of linear least squares fitting was relatively low comparing to other parameters. Although, one can hardly recognize a clear pattern in the plotted points:

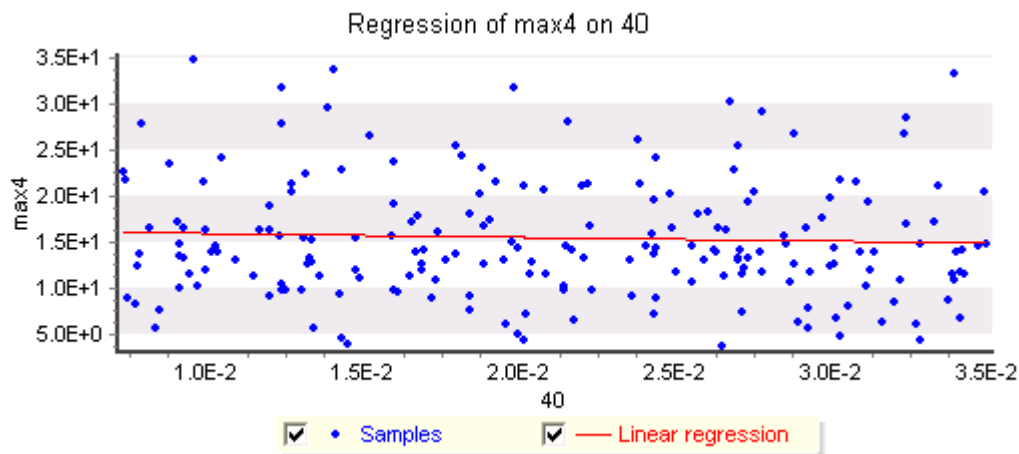


Figure 5.19 Values of phosphorus to carbon ratio of Flagellates type P (40) plotted against the maximal chlorophyll-a concentrations computed for these points at station Schouwen10. The red line was fitted using linear least squares

Further, the potential correlations between 40 and remaining parameters were investigated. Unfortunately, no significant association of the considered parameters was observed. One possible explanation of this fact might be that the product moment correlation measures only the degree of linear relationship between variables, when the rank correlation is used to check the monotone relationship. In case of our model, the correlations between the factors, strongly depending on the values of other factors, might be much more complicated. It can be concluded that no major monotone relationships are engaged by the model (rank correlations are not higher than 0.1 in absolute value). The following figure shows the obtained fit to conditional expectation of the maximal chlorophyll-a concentration when the value of 40 is known:

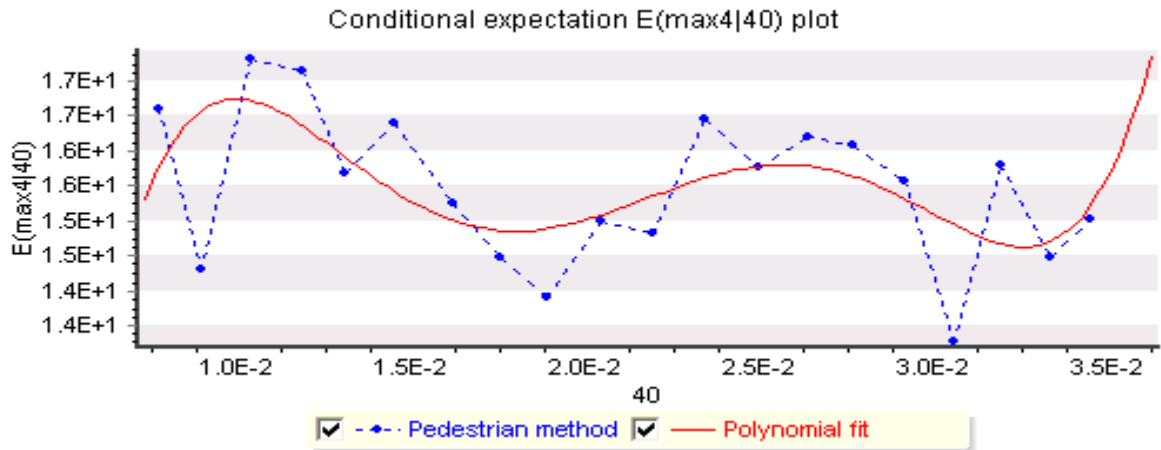


Figure 5.20 Conditional expectation of the maximal chlorophyll-a concentration at station Schouwen10 given the values of P:C ratio of Flagellates type P (40) computed with use of the Pedestrian method and polynomial fit (degree 5)

A polynomial of degree 5 was used. Shape of the above curve explains a low value of the corresponding correlation ratio. This feature was repeated across the results generally, with exclusion of 9 locations, providing low sensitivities due to change in P:C ratio of Flagellates type P and for more than a half of the considered stations low sensitivities due to change in P:C ratio of Flagellates type E. For comparison purpose, the same plot made for Chlorophyll-a to carbon ratio of algae type E is shown below:

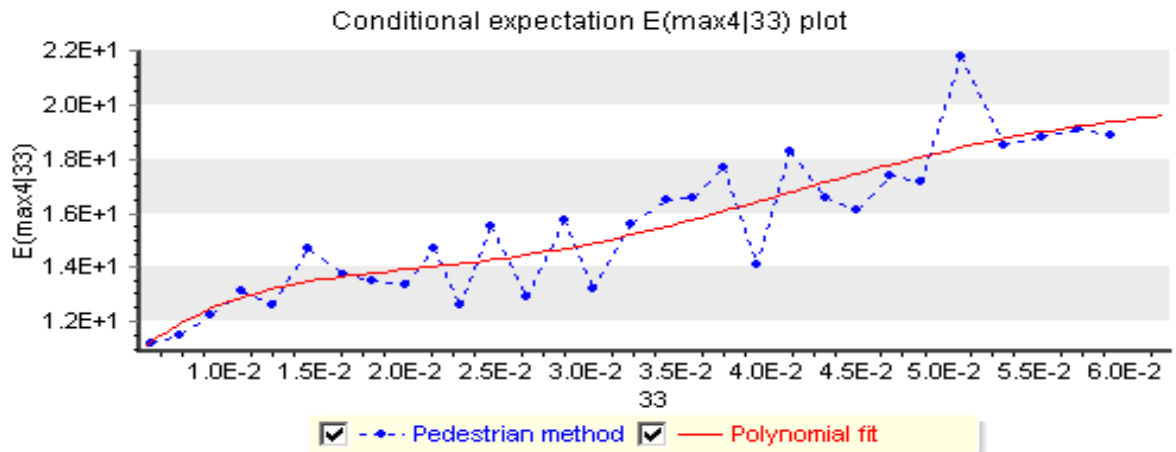


Figure 5.21 Conditional expectation of the maximal chlorophyll-a concentration at station Schouwen10 given the values of Chlfa:C ratio of type E (33) computed with use of the Pedestrian method and polynomial fit (degree 5)

Evidently, there is a trend in the plotted conditional expectation of the maximal concentration. Increasing the value of parameter 33 leads to significant increase of the average size of the

bloom, hence the variation in Equation (13) must also be high. Therefore, as it can be observed in Figure 5.18, the correlation ratio computed for this factor got the greatest value among the parameters.

With respect to growth rates, two applied methods seem to be on the contrary to each other. According to the Morris approach, at Schouwen10 parameters 60, 62, 63, 65 and 69 are of major importance. However, the variance analysis indicated that only growth rate of *Phaeocystis* type E (69) is still leading but the others fell below the level of 1 % contribution. Especially for growth rate of Diatoms type E (60) the correlation ratio is considerably low of around 0.2 %. On the other hand, growth rate of Diatoms type N (61), that got a rank of 36 with respect to the Euclidean distance, appeared to have a quite strong influence on the output according to not only the correlations ratio but also rank and product moment correlations. Parameter 69 was indicated as a commonly influential input factor. A bit less important is 61 at more than a half of the selected stations, but not at for example distant locations along Terschelling transect or stations close to Terheijde1. The greatest impact of this parameter can be observed at Appelzak1 and Appelzak50:

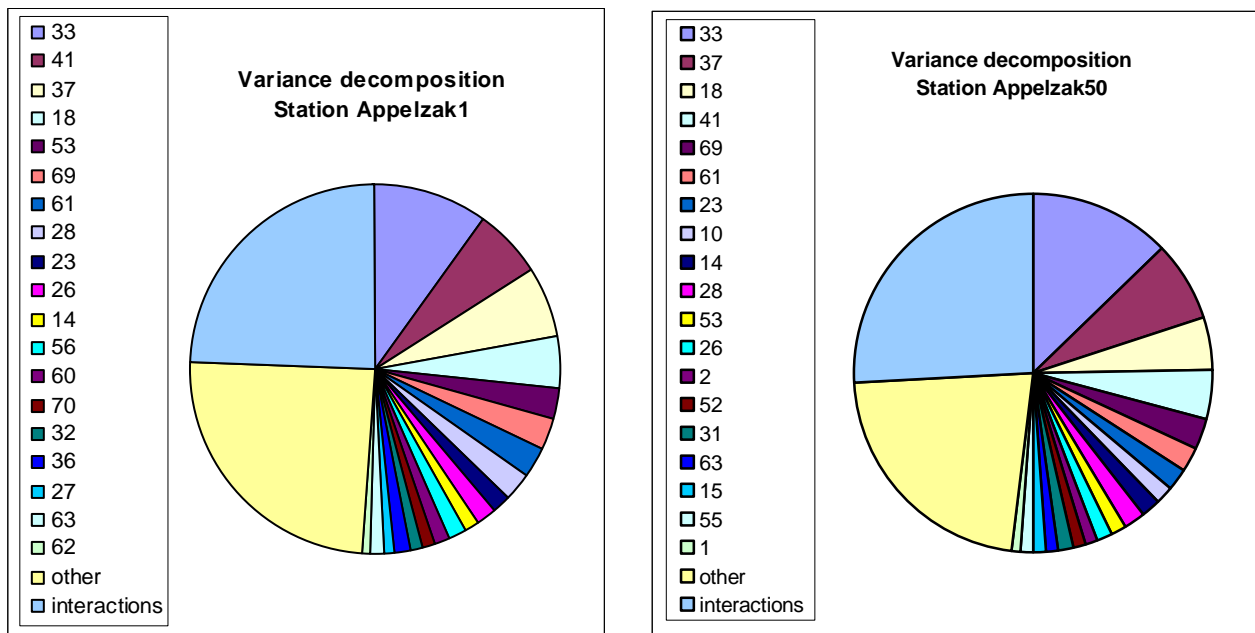


Figure 5.22 Variance decomposition of the output at Appelzak1 and Appelzak 50

Not much can be said about 62 and 63, except the fact that influence of the first one increases for instance along Callantsoog transect, Terschelling004 and Terschelling010 that are both also the case for the second one including also Terheijde transect. Growth rates of other types of

Flagellates appeared to be fairly negligible. Interestingly, this method detected the influence of Dino-Flagellates and some parameters related to this species, in particular growth rate of type P (68) affected the maximal bloom at some locations like distant Noordwijk, Terheijde and Egmond situated in the same area of the North Sea. Comparing to Phaeocystis type E, growth rates of other types are less important – again only for a small range of stations the contribution of these parameters was greater than 1 %.

Study of the results revealed that P:C ratio of Phaeocystis type E (53) fairly influenced the output at almost all the considered locations. Moreover, other parameters corresponding to this species seem to have significant effects on the maximal chlorophyll-a concentration but only at particular stations. For instance, N:C ratios of all types of Phaeocystis are of high importance along Terschelling transect. Specific extinctions were found as quite negligible, with the exception of type E (51) that at few locations brought the contribution to variance decomposition greater than 1 %, for example at Goeree06, situated next to it Terheijde1 and a bit to the east Noordwijk02 and Noordwijk10.

In the Figure 5.18, one can notice an increase of importance of settling velocity of algae. Parameter 10 corresponding to Flagellates type P appeared among the most influential factors. However, this pattern was repeated over the stations for only this type of algae. Even more surprising is the fact that specific extinction of Dino-Flagellates type P (48), with a rank of 68! in the Morris method, was pointed out as fairly important not only at this station. Another parameter corresponding to this particular type that could be seen as meaningful is N:C ratio (49). It is showed in the following figure made for Egmond1:

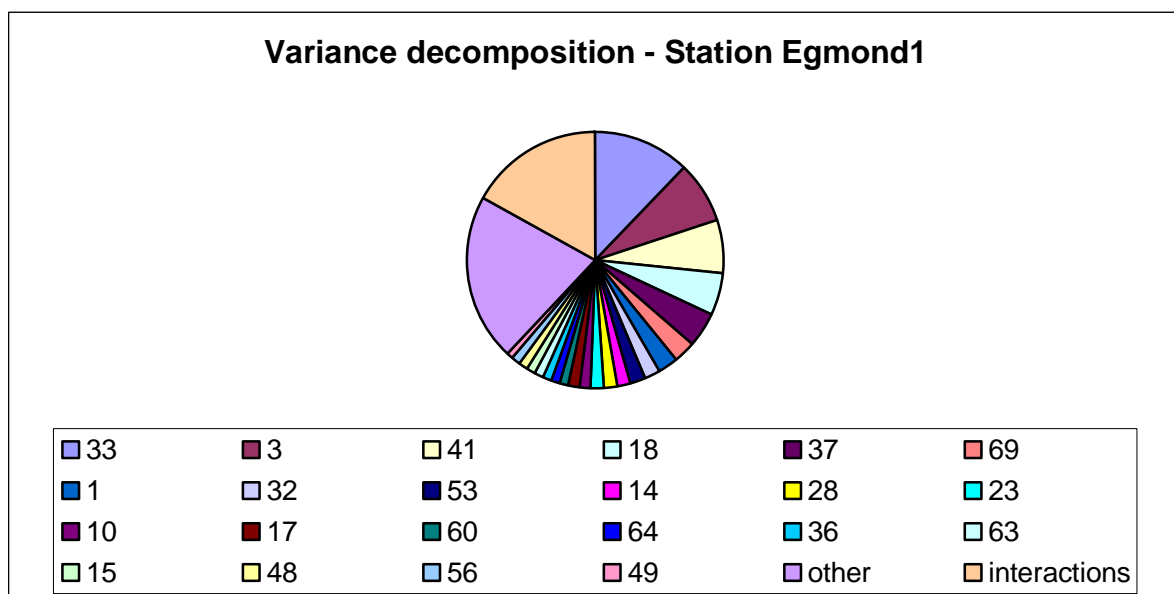


Figure 5.23 Variance decomposition of the output at station Egmond1

Other specific extinctions of Dino-Flagellates as well as other parameters of this species were found as more negligible throughout the considered area.

Returning to Figure 5.18, another phenomenon is the appearance of specific extinction of Flagellates type E (30) in the group of high importance that achieved the highest value of the correlation ratio at this particular location namely 1.16 %. In the plot of elementary effects it is laying relatively close to the point (0,0), that can be observed in Figure 5.2. Taking into account all specific extinctions only one seems to be repeated as influential over a great number of stations, namely of Diatoms type E (15) but again not on a large scale. The last relevant parameter indicated by Figure 5.18 is N:C ratio of Diatoms type N (26). Briefly, its impact was also noticeable at different places, with exclusion of Noordwijk where it was observed to be relatively weak, and 26 can be considered as one of important parameters in the GEM model. In the same way as it was pointed out by the elementary effects method, this parameter together with other N:C ratios, is of high importance especially at Terschelling175 [App].

It is worth mentioning that for the last four discussed factors (10, 48, 30 and 26) the values of rank and product moment correlations with the maximal chlorophyll-a concentration as well as the values of regression coefficients were smaller comparing to the other parameters found as important, even if calculated correlation ratios were almost the same. Hence, it is clear that with a different measure of importance one can determine a different order of parameters.

Obviously, over distinct locations the variance decomposition of the maximal chlorophyll-a concentration was changed, due to various characteristics of stations. Similarly to the results produced by the Morris method, P:C ratio of Diatoms type E (17) was detected as one of the most important parameters but only for a half of considered locations, especially in the area between Goeree06 and Terschelling close to the shore. In the same manner, P:C ratio of Diatoms type P (22) was confirmed to be influential at the same places as were found by the previous approach, in particular Noordwijk10, Noordwijk30 and Egmond10 known to be places of high concentration of Diatoms. The plots are shown in Appendix and one can also notice there the presence of other parameters related to this species. Another interesting observation is that parameter 31 corresponding to N:C ratio of Flagellates type E turned out to be fairly negligible, but still its influence is emphasized at Terschelling. Moreover, N:C ratio of Flagellates type N (35) was assigned a very low value of the correlation ratio at all the considered stations, namely less than 0.6 %, and it can be concluded that its importance decreased remarkably comparing to the results of previous approach. Evidently, it is hard to generalize the outcome of the variance-based analysis due to specific local characteristics of

the considered stations. A penetrating analysis of the computed values of correlation ratio is needed for each location separately to have a comprehensive view of the influence of parameters on the chlorophyll-a concentration.

4.4 Variance-based analysis - annual average chlorophyll-a concentration.

Finally, let us focus on the annual average concentration of chlorophyll-a. Exactly as in the previous case the analysis is first conducted for station Schouwen10. Below the variance decomposition of the output is presented:

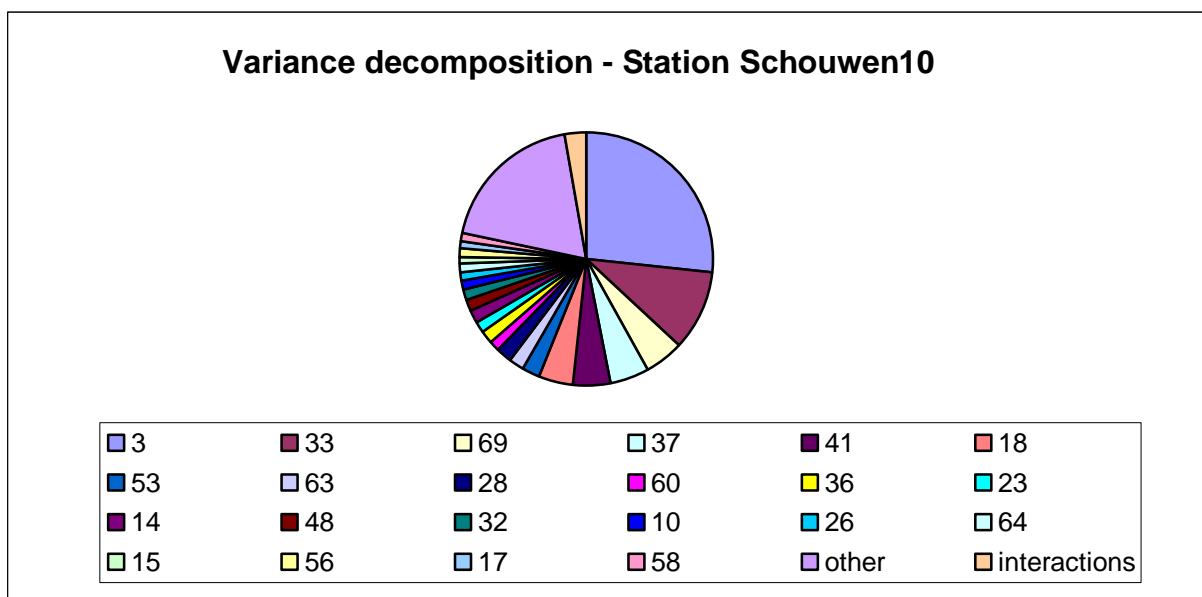


Figure 5.24 Variance decomposition of the output at station Schouwen10

In the same manner as in the Morris method the importance of specific extinction of inorganic matter (3) increased remarkably for the new output function. It is the case not only for this particular location. In general, it is considered as very influential for all the selected stations and the maximal correlation ratio computed for this parameter jumped up to 26.7 % while for the maximal concentration it was only 12 % and it reached the threshold of 1 % for only half of the analyzed locations. Other parameters related to light do not provide a high contribution to the output variance at Schouwen10. Once again, for the locations mentioned in the analysis of the previous output, specific extinction due to humic substances (1) is of major importance and at the same stations as considered previously parameter 2 makes more or less the same contribution. Parameter 4 corresponding to extinction of dead particulate matter (4) can be

assumed to be fairly negligible because of the fact that its correlation ratio for any location did not exceed the value of 0.8 %.

Chlorophyll-a to carbon ratios preserved their ranks of importance also for the annual average concentration. Although, the ones corresponding to Diatoms have now a weaker effect on the output, for instance the maximal correlation ratio of parameter 18 decreased from 11 % down to 8 % at Noordwijk30.

With respect to settling velocities the one of detritus (14) should be considered as a commonly important parameter in the model but again to the lesser degree. Additionally, at Schouwen10 and at almost all the stations settling velocity of Flagellates type P (10) was found somewhat influential. Moreover, at station Dantziggat located a bit to the east from Terschelling the one corresponding to Phaeocystis type P (13) made a quite remarkable contribution to the output variance for both the maximal and average concentration:

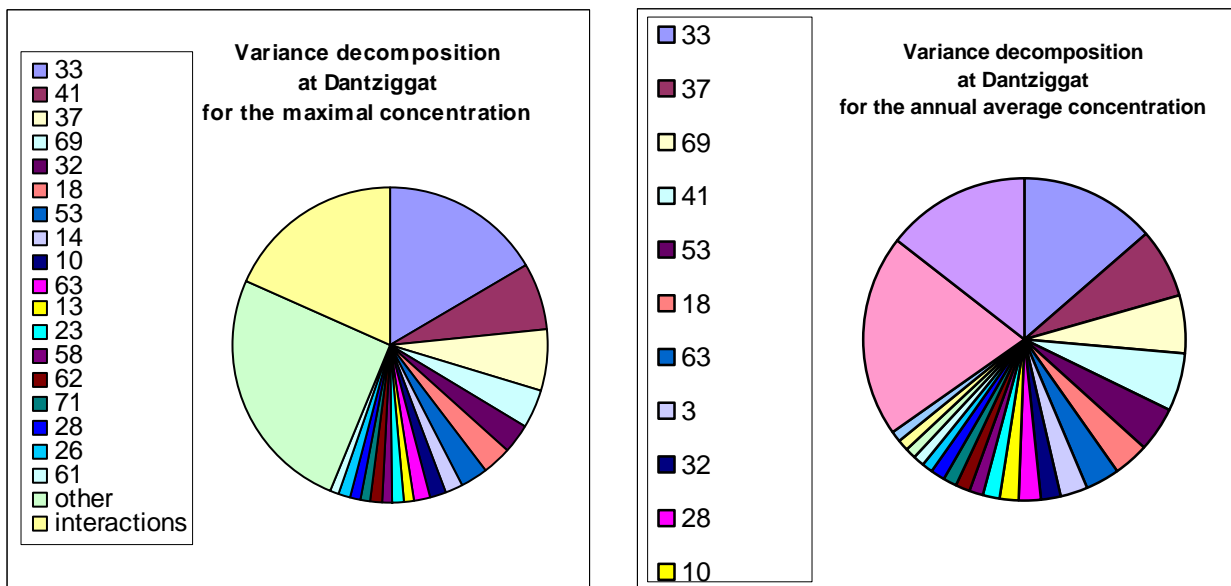


Figure 5.25 Variance decomposition of the output (on left the maximal chlorophyll-a concentration, on right the annual average) at station Dantziggat

But that is the only location where its contribution was bigger than 1 %. The effects of changes in other settling velocity values did not affect the output variance strongly.

Moreover, as it showed by Figure 5.24 factor 32 has a relatively strong effect on the average concentration, in particular at Schouwen10. The change of the output function did not affect the strength of its influence that is still observed at the same locations forming more than a half of analyzed stations. Once again, parameter 40 did not appear among the most influential factors (Figure 5.24). Its contribution to the output variance was smaller than 0.5 % at this location. No significant relationship with the output function was detected neither by the

product moment correlation nor rank correlation. Additionally, the computed regression coefficient did not plant any doubts about its negligible effect at this location.

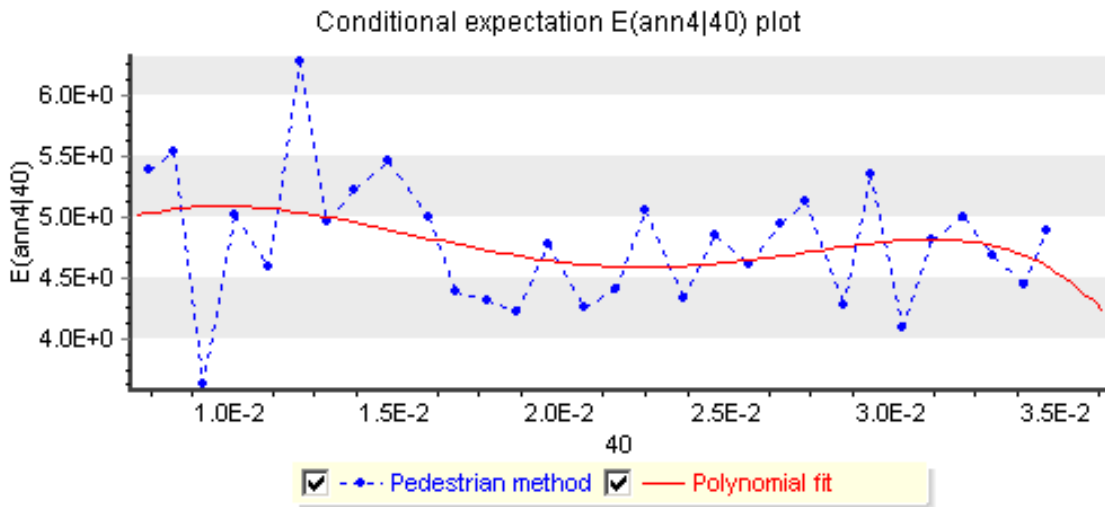


Figure 5.26 Conditional expectation of the annual average chlorophyll-a concentration at station Schouwen10 given the values of P:C ratio of Flagellates type P (40) computed with use the Pedestrian method and polynomial fit (degree 5)

The pattern was repeated over the stations yet again showing however quite low values of the linear regression coefficients that could indicate a kind of strong influence on the output and on the other hand low values of the correlation ratios suggesting no importance of this parameter.

Admittedly, in the Figure 5.24 growth rates of algae have caused a different composition of the variance contribution comparing to the Figure 5.18. Clearly, the one of Phaeocystis type E is the leader once again and the minimal correlation ratio computed for this parameter, at Terschelling235, is equal to 2.15 % attaching great importance to this factor. In general, the influence of parameters 61, 62, 68, 70 and 71 weakened slightly. For 61 the value of correlation ratio at Appenzak1 is now around 1 %. To study the behaviour of these parameters more carefully one can have a look at the tables of results [13]. Considering other types of Dino-Flagellates (66 and 67), they turned out to be fairly negligible exactly as in the previous case. On the other hand, the influence of growth rates of Diatoms type E and Flagellates type E (60 and 63) has increased remarkably, that is expressed in a greater number of station where these two parameters were pointed out as ones of the most influential factors and higher maximal values of the correlation ratio. In Figure 5.24 one can notice that parameter 64 was included as well in the list of parameters making significant contributions to the variance of the

output, however comparing the size of its contribution to the others it is still relatively small and this fact was also confirmed at different stations.

Changes in single extinctions of algae do not affect the output strongly for both the maximal and average concentration. However, there are few exceptions to this observation, namely specific extinction of Diatoms type E (15) in both cases was indicated as important for more than a half locations including Terschelling, Calandsoog and Terheijde, but still on a small scale (less than 2 %). Parameter 51 standing for extinction of Flagellates type E reached the level of 1 % only at one of the previously mentioned locations that is Terheijde1. Interestingly, with this new output function it turned out that specific extinctions of Dino-Flagellates type N and P (45 and 48) should be seen as influential. The first one was fairly unimportant for the maximal concentration and the latter seemed to matter at only a few stations, while now it is the case for around 30 locations. The significance of factor 45 was easily noticed in the Wadden Sea. Further, one will observe an increase of importance of this species expressed also in higher values of correlation ratios computed for other parameters and this fact corresponds well with our intuition.

A few general statements can be made about N:C and P:C ratios of algae. Firstly, as it was mentioned before, not only specific extinction of Dino-Flagellates type P influences the annual average evidently, but also N:C and P:C ratio of this particular type (49 and 50). The latter mostly at the stations situated around Noordwijk70 and the first one at Egmond and Terheijde closer to the shore but both with only a little contribution. Other characteristics of this species, excluding specific extinction of type N, can be considered as negligible. Among P:C ratios of Diatoms (17, 22 and 27) only the first one corresponding to type E should be seen as important especially in the area mentioned previously, however its effect has weakened (the maximal value of the correlation ratio dropped from 4.5 % to 2.4 % at Harlingen). For Flagellates type N the importance of P:C ratio (36) was extended to almost all the stations, but at the low level. Briefly, parameters 53 and 56 have established their significance also for the annual average and N:C ratios specified for this species (52, 55 and 58) strengthen their importance particularly along Terschelling transect. Also there N:C ratios of Diatoms, as well at Schouwen10 for type P, and N:C ratio of Flagellates but only type E preserved their rank of importance. In the same way as in the Morris method Si:C ratios turned out to be unimportant particularly for the annual average chlorophyll-a concentration, with respect to all the used measures.

5 Discussion

The Generic Ecological Model is an instrument that can be applied to any water system (fresh, transitional or coastal) to calculate the primary production, chlorophyll-a concentration and phytoplankton species composition. It includes a consistent set of formulations of processes, that together describe a part of the ecosystem functioning. It produces sufficiently accurate results in a reasonable time. It has been used as the basis for several major policy and management decisions.

Sensitivity analysis is a fundamental tool in the construction, use and understanding of models. It may identify the most important factors within a model and check if the model resembles the system under study. The obtained results can be used to fix unimportant factors. Furthermore, they may prioritize further research and experiments addressing the estimates of those parameters that have the greatest effect on the output of interest. Moreover, the information obtained can be used to improve satellite data assimilation by estimating significant parameters. The model was analyzed to assess the sensitivity of a subset of the model outputs, to a subset of the input parameters. Sensitivity analysis of GEM was conducted for 71 ecologically significant parameters concerning light and algae's characteristics and for two output functions, namely the maximal and the annual average concentration. Additionally, the change in the model response was analyzed at 49 monitoring stations representing the diversity of characteristics of the Dutch part of the North Sea.

Not many sensitivity analysis methods could be considered as suitable for the analysis of GEM, due to several reasons including the model complexity, non-linearity, engagement of a great number of interacting parameters and, the most important, a relatively long time of a single simulation. Hence, the selected group of appropriate methods focused on global approaches and on OAT (one at a time) designs, but only on those that overcome the limitations of the local approach. Finally, it was decided to use the Morris method and later the analysis was enriched by applying the variance based methods. To avoid complexity of computations the total sensitivity indices were not calculated.

The obtained results at different locations correspond well with local water conditions. Commonly, chlorophyll-a to carbon ratios were found significant for both output functions. Moreover, comparing Tables 5.2 and 5.3 one can notice that phosphorus to carbon ratios of Flagellates type E and P, maximum growth rates of Diatoms, Flagellates and Phaeocystis type E and of Diatoms type P and additionally extinction of inorganic suspended matter are of crucial importance in both cases. These conclusions agree with expert knowledge of the

ecological processes in the North Sea [4]. With respect to the maximal concentration, the light related parameters, excluding extinction of dead particulate matter, turned out to be quite significant. The annual average concentration was less influenced by the changes in background extinction and extinction due to humic substances in pure fresh water. Furthermore, the analysis showed that *Phaeocystis* is a species that has a long-run impact on the chlorophyll-a concentration. Settling velocities of algae and characteristics of *Dino-Flagellates*, but not of type P, can be considered as fairly insignificant. For other parameters a penetrating analysis of the computed values of correlation ratio is needed for each location separately, in order to have a comprehensive view of their influence on the chlorophyll-a concentration.

It is worth mentioning that the factor-screening methods, in particular the Morris approach, are crucial when developing metamodels. Details can be found in [10].

As it was mentioned before, time of blooming was also considered as a possible output, but some combinations of the values of parameters produced unrealistic model responses. The generated Monte Carlo sample, that was processed in the variance-based analysis, can also be used to narrow down the overestimated ranges of the parameters by checking, for example by cobweb plots in Unicorn, which combinations caused the improbable shift of the blooming time. Although, one should be aware of the difficulty of this task caused by existing interactions and moderately large number of parameters.

The effect of interactions between parameters on the chlorophyll-a concentration was not investigated. One of the methods that can be applied to evaluate this effect is the New Morris method. This technique designs a second set of pathways so as to estimate the second-order effects, or second derivatives, of the output with respect to the input parameters. This second-order sensitivity analysis provides a mean value and a corresponding standard deviation, for each pairwise interaction between factors for a given output, where the samples are taken over the parameter space. Both methods, the Morris and the New Morris approach, are based in graph theory, and use the optimal number of model evaluations to reduce computation time, for a given accuracy. More information can be found in [8]. Interestingly, the special software has been developed by Campolongo and Braddock at Griffith University (QLD, Australia) and it is an open source [9]. For the analysis of GEM, the method has not been applied due to the time constraints and possible difficulties in analyzing the second-order effects for a great number of input parameters, in particular for 71 it leads to the study of around 2.500 pairwise interactions for 49 stations for two output functions. Another approach that could be applied to

check the effect of interactions is the computation of total sensitivity indices, i.e. the extended Fourier Amplitude Sensitivity Test.

The variance-based method that was used in the analysis can be improved for example by Latin hypercube sampling, that allows covering the whole input space or by taking a larger sample. The Morris method can produce better estimates if one generates a high number of different trajectories and then chooses only those which maximize their spread in the input space. This procedure allows to explore the input space thoroughly.

Methods used in the research could be also applied in case of non-uniform distributions of the considered input parameters, in particular truncated normal distributions that can be sometimes found in the literature. Sensitivity analysis results generally depend more on the selected ranges than on the assigned distributions, hence in the analysis all the parameters were assumed to be uniformly distributed.

It is possible to conduct sensitivity analysis for separate modules within the ecological part of GEM, i.e. suspended sediment, to have a greater choice of methods at hand. However, it may lead to a dangerously incomplete exploration of the uncertainties, i.e. interactions can be overlooked. Making analysis of the entire model, together with hydrodynamics, would create some difficulties, in particular with a huge number of factors to analyze, since only the ecological module uses more than four hundreds of parameters. Moreover, the ranges for all the parameters cannot be found in the literature and in the first place distributions should be derived from the data or estimated with use of expert judgement.

Sensitivity analysis is closely linked to uncertainty analysis, which quantifies the overall uncertainty in the model output as a result of uncertainties in the model input. The model evaluations that had to be run for the sensitivity analysis can be also used to estimate means, standard deviations, confidence bounds and cumulative distribution functions of the response variables.

References

- 1) Anouk N. Blauw, Hans F.J. Los, Marinus Bokhorst and Paul L.A. Erfemeijer, 2008, GEM: a Generic Ecological Model for estuaries and coastal waters.
- 2) Hans F.J. Los, 1991, Mathematical Simulation of algae blooms by the model BLOOM II.
- 3) Delft3D-WAQ User Manual.
- 4) F.J. Los, M.T. Villars and M.W.M van der Tol, A 3-dimensional primary production model (BLOOM/GEM) and its application to the (southern) North Sea.
- 5) <http://dutiosc.twi.tudelft.nl/~risk> and then software/Unicorn_or directly [download](#)
- 6) A.Saltelli, M.Ratto, T.Andres, F.Campolongo, J.Cariboni, D.Gatelli, M.Saisana, S.Tarantola, 2008, Global Sensitivity Analysis, the Primer.
- 7) A.Saltelli, 2004, Sensitivity analysis in practice: a guide to assessing scientific models.
- 8) R. Cropp and R. Braddock, 2002, Reliability Engineering and System Safety 78, page 77-83
- 9) The person to be contacted is Francesca Campolongo, currently at: Institute for System, Informatics and Safety, TP 361, Joint Research Centre of the European Communities, 21020 Ispra, Varese, Italy; Tel.: +39-332-785476; Fax: +39-332-785733; E-mail: francesca.campolongo@jrc.it.
- 10) F.M. Alam, K.R. McNaught, T.J. Ringrose, 2004, Using Morris' randomized OAT design as a factor screening method for developing simulation Metamodels.
- 11) R.Cooke, J. van Noortwijk, 1998, Local probabilistic sensitivity measures for comparing FORM and Monte Carlo calculations illustrated with dike ring reliability calculations.
- 12) D.C. Montgomery, 2001, Design and analysis of experiments
- 13) http://rapidshare.de/files/40226030/Results_Sensitivity_Analysis_of_GEM.doc.html

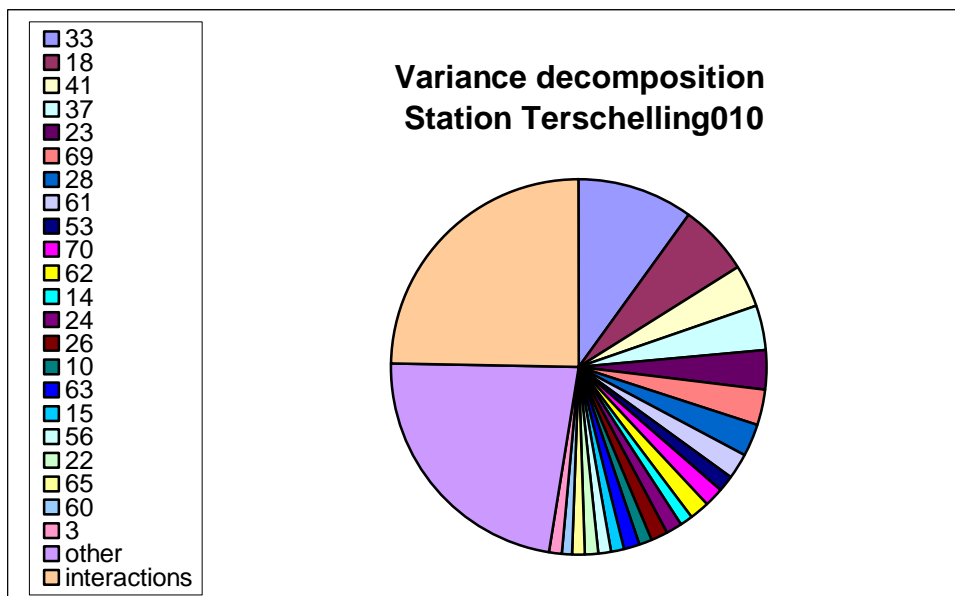
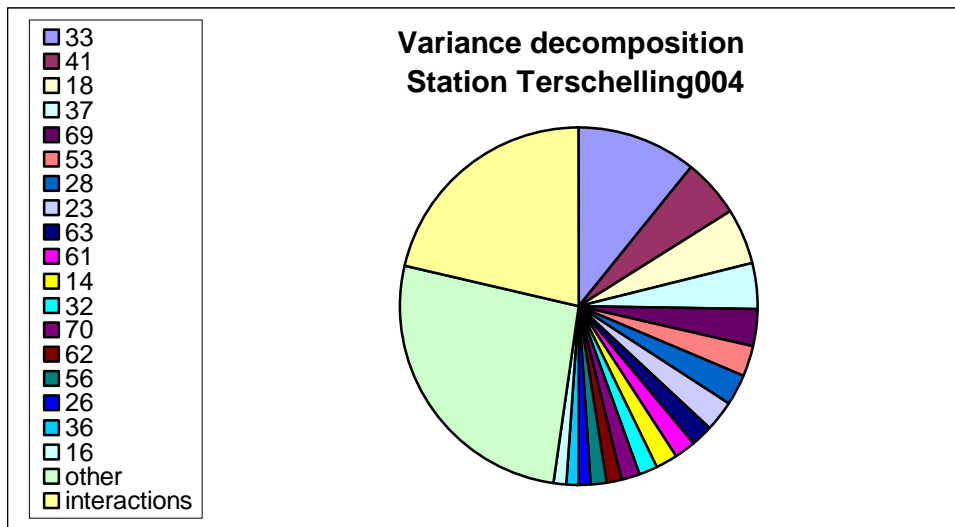
6 Appendix

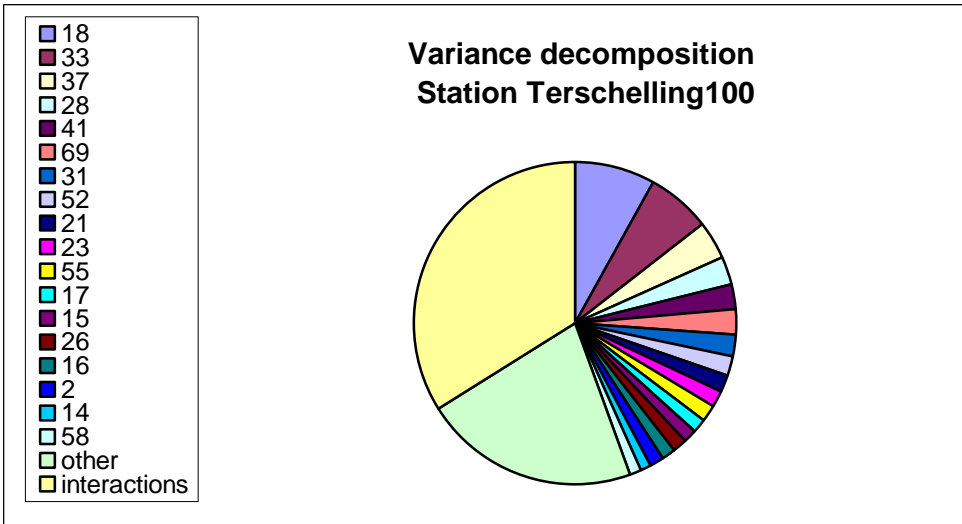
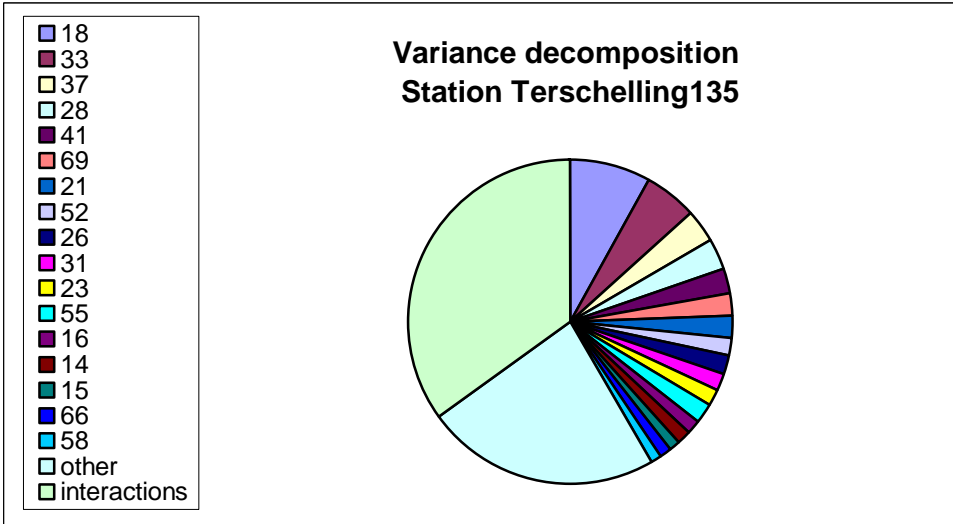
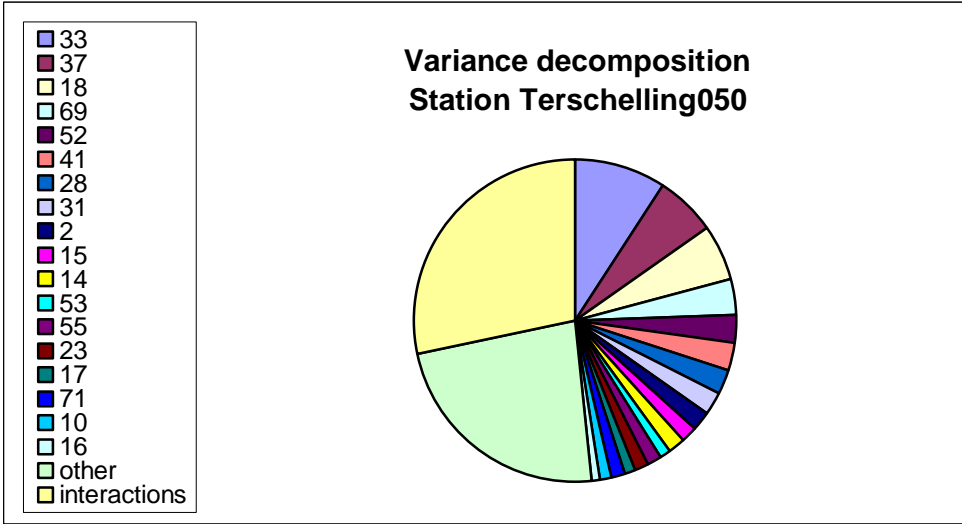
The list of parameters and their ranges:

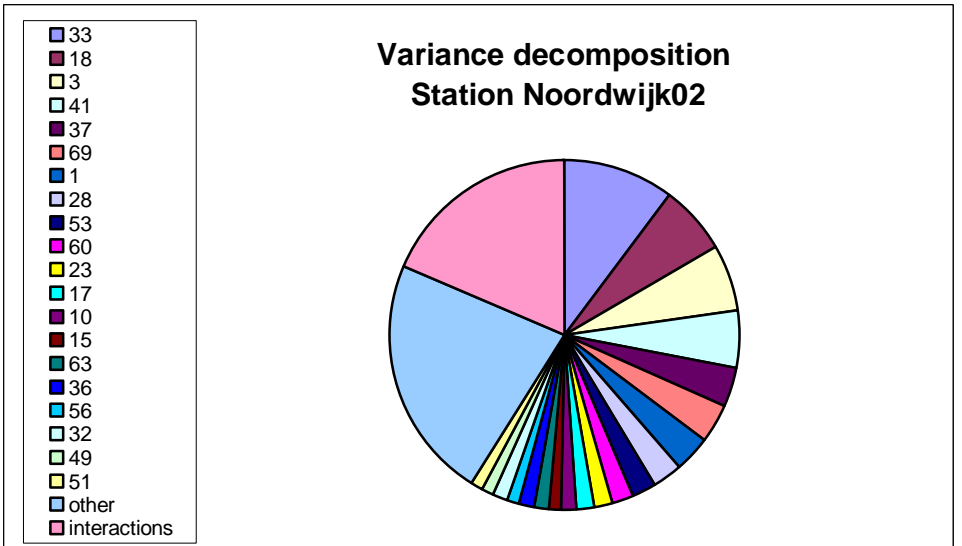
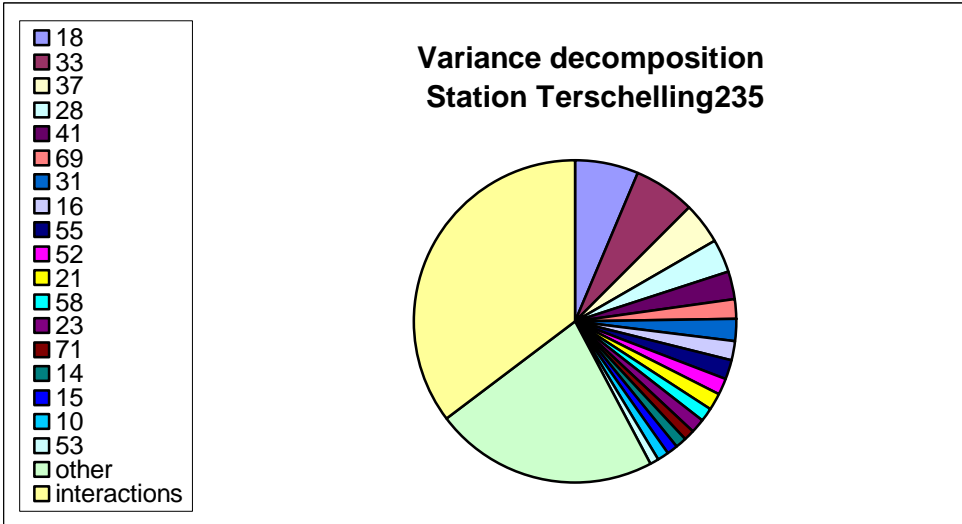
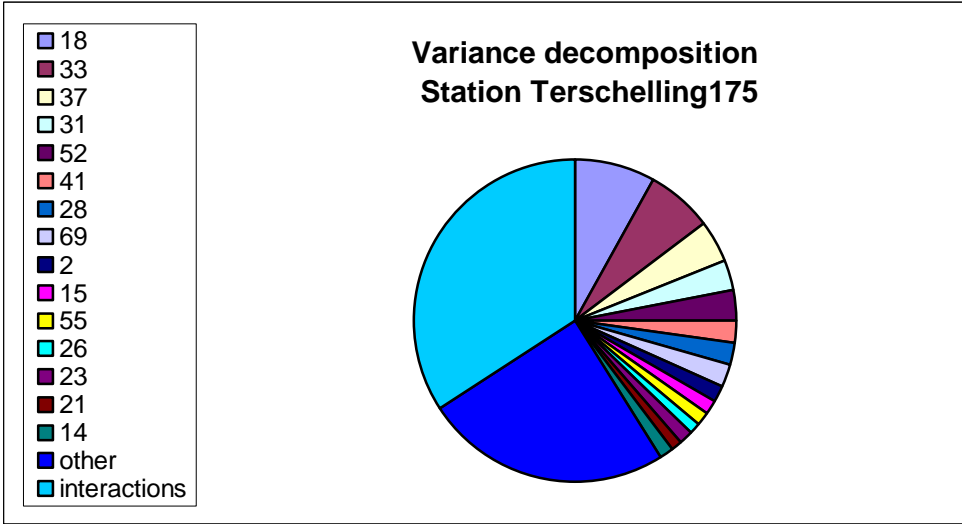
No	Parameter	Description	range_fr	range_to	base
1	$E_{Hum,0}$	extinction due to humic substances	0.50000	1.00000	0.97000
2	K_b	background extinction	0.04000	0.08000	0.08000
3	E_{SPM}	specific extinction of inorganic suspended matter	0.01000	0.05000	0.02500
4	E_{POM}	specific extinction of particulate dead organic matter	0.10000	0.20000	0.10000
5	$V_i, i=1, \dots, 9$	settling velocity for algae Diatoms type E	0.25000	0.75000	0.50000
6		settling velocity for algae Diatoms type N	0.50000	1.50000	1.00000
7		settling velocity for algae Diatoms type P	0.50000	1.50000	1.00000
8		settling velocity for algae Flagellates type E	0.00000	0.25000	0.00000
9		settling velocity for algae Flagellates type N	0.25000	0.75000	0.50000
10		settling velocity for algae Flagellates type P	0.25000	0.75000	0.50000
11		settling velocity for algae Phaeocystis type E	0.00000	0.25000	0.00000
12		settling velocity for algae Phaeocystis type N	0.25000	0.75000	0.50000
13		settling velocity for algae Phaeocystis type P	0.25000	0.75000	0.50000
14	V_{POM}	settling velocity of particulate dead organic matter	1.00000	2.00000	1.50000
15	E_1	specific extinction of algae Diatoms type E	0.15000	0.45000	0.24000
16	$S_{N,1}$	N:C ratio for algae Diatoms type E	0.05000	0.30000	0.25500
17	$S_{P,1}$	P:C ratio for algae Diatoms type E	0.01000	0.03500	0.03150
18	$S_{chl,1}$	chlorophyll-a:C ratio for algae Diatoms type E	0.00500	0.06000	0.05330
19	$S_{Si,1}$	Si:C ratio for algae Diatoms type E	0.15000	0.55000	0.44700
20	E_2	specific extinction of algae Diatoms type N	0.15000	0.45000	0.21000
21	$S_{N,2}$	N:C ratio for algae Diatoms type N	0.05000	0.30000	0.07000
22	$S_{P,2}$	P:C ratio for algae Diatoms type N	0.01000	0.03500	0.01200
23	$S_{chl,2}$	chlorophyll-a:C ratio for algae Diatoms type N	0.00500	0.06000	0.01000
24	$S_{Si,2}$	Si:C ratio for algae Diatoms type N	0.15000	0.55000	0.28300
25	E_3	specific extinction of algae Diatoms type P	0.15000	0.45000	0.21000
26	$S_{N,3}$	N:C ratio for algae Diatoms type P	0.05000	0.30000	0.10500
27	$S_{P,3}$	P:C ratio for algae Diatoms type P	0.00960	0.03500	0.00960
28	$S_{chl,3}$	chlorophyll-a:C ratio for algae Diatoms type P	0.00500	0.06000	0.01000
29	$S_{Si,3}$	Si:C ratio for algae Diatoms type P	0.15000	0.55000	0.15200
30	E_4	specific extinction of algae Flagellates type E	0.15000	0.45000	0.25000
31	$S_{N,4}$	N:C ratio for algae Flagellates type E	0.05000	0.30000	0.20000
32	$S_{P,4}$	P:C ratio for algae Flagellates type E	0.01000	0.03500	0.02000
33	$S_{chl,4,7,10}$	chlorophyll-a:C ratio for algae type E	0.00500	0.06000	0.02280
34	E_5	specific extinction of algae Flagellates type N	0.15000	0.45000	0.22500
35	$S_{N,5}$	N:C ratio for algae Flagellates type N	0.05000	0.30000	0.07750
36	$S_{P,5}$	P:C ratio for algae Flagellates type N	0.00960	0.03500	0.00960
37	$S_{chl,5,8,11}$	chlorophyll-a:C ratio for algae type N	0.00500	0.06000	0.00670

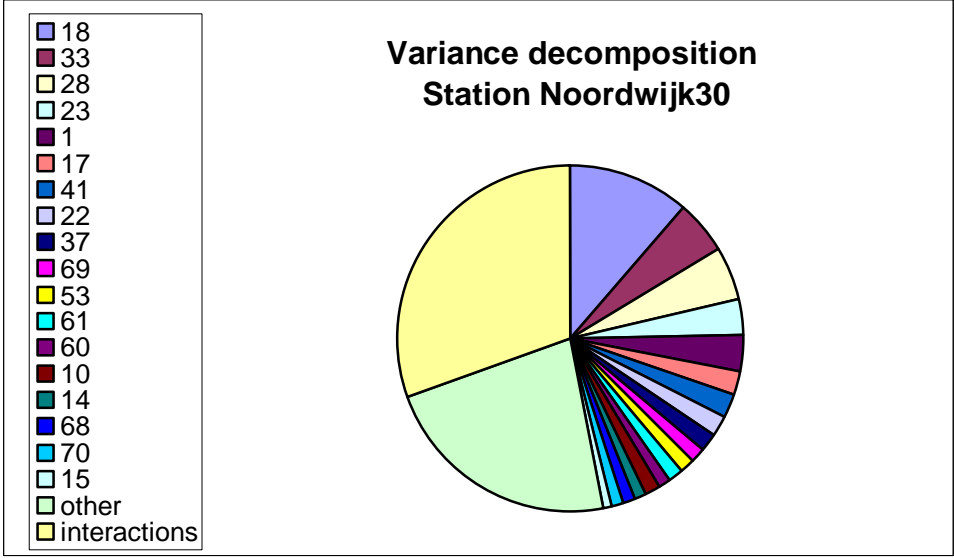
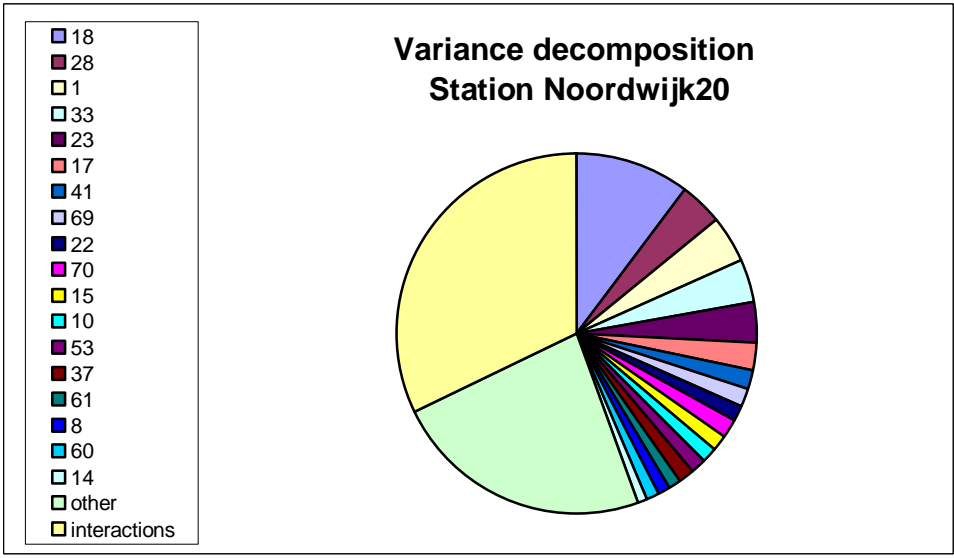
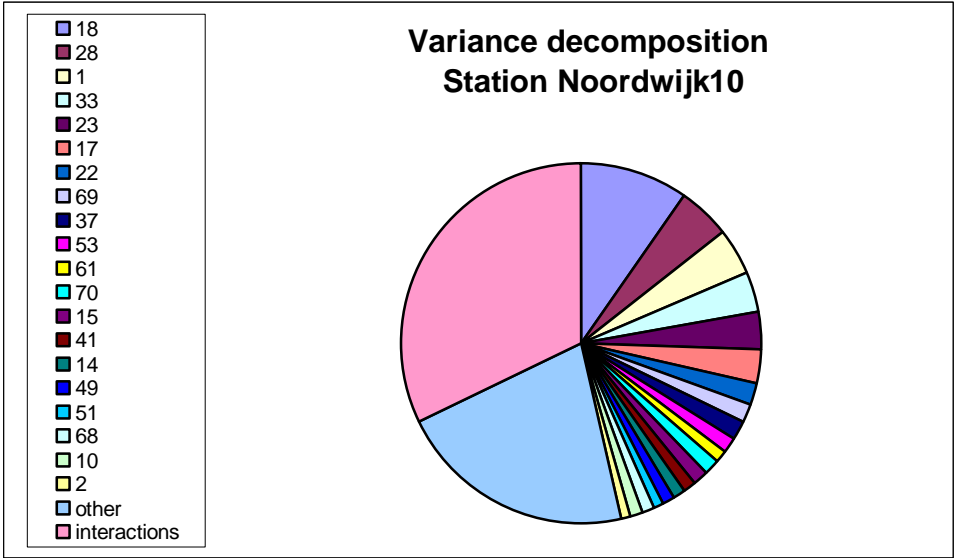
38	E_6	specific extinction of algae Flagellates type P	0.15000	0.45000	0.22500
39	$S_{N,6}$	N:C ratio for algae Flagalletas type P	0.05000	0.30000	0.11250
40	$S_{P,6}$	P:C ratio for algae Flagalletas type P	0.00720	0.03500	0.00720
41	$S_{chl,6,9,12}$	chlorophyll-a:C ratio for algae type P	0.00500	0.06000	0.00670
42	E_7	specific extinction of algae DinoFlagellates type E	0.15000	0.45000	0.20000
43	$S_{N,7}$	N:C ratio for algae DinoFlagellates type E	0.05000	0.30000	0.16250
44	$S_{P,7}$	P:C ratio for algae DinoFlagellates type E	0.01000	0.03500	0.01675
45	E_8	specific extinction of algae DinoFlagellates type N	0.15000	0.45000	0.17500
46	$S_{N,8}$	N:C ratio for algae DinoFlagellates type N	0.05000	0.30000	0.06388
47	$S_{P,8}$	P:C ratio for algae DinoFlagellates type N	0.01000	0.03500	0.01120
48	E_9	specific extinction of algae DinoFlagellates type P	0.15000	0.45000	0.17500
49	$S_{N,9}$	N:C ratio for algae DinoFlagellates type P	0.05000	0.30000	0.07088
50	$S_{P,9}$	P:C ratio for algae DinoFlagellates type P	0.00960	0.03500	0.00960
51	E_{10}	specific extinction of algae Phaeocystis type E	0.15000	0.45000	0.45000
52	$S_{N,10}$	N:C ratio for algae Phaeocystis type E	0.05000	0.30000	0.18750
53	$S_{P,9}$	P:C ratio for algae Phaeocystis type E	0.01000	0.03500	0.02250
54	E_{11}	specific extinction of algae Phaeocystis type N	0.15000	0.45000	0.41300
55	$S_{N,11}$	N:C ratio for algae Phaeocystis type N	0.05000	0.30000	0.07500
56	$S_{P,11}$	P:C ratio for algae Phaeocystis type N	0.01000	0.03500	0.01360
57	E_{12}	specific extinction of algae Phaeocystis type P	0.15000	0.45000	0.41300
58	$S_{N,12}$	N:C ratio for algae Phaeocystis type P	0.05000	0.30000	0.10375
59	$S_{P,12}$	P:C ratio for algae Phaeocystis type P	0.01000	0.03500	0.01056
60	$P_{i,0}$ for $i=1, \dots, 12$	maximum growth rate at 0 °C for Diatoms type E	0.05000	0.15000	0.07000
61		maximum growth rate at 0 °C for Diatoms type N	0.05000	0.15000	0.05400
62		maximum growth rate at 0 °C for Diatoms type P	0.05000	0.15000	0.05400
63		maximum growth rate at 0 °C for Flagalletas type E	0.05000	0.15000	0.09000
64		maximum growth rate at 0 °C for Flagalletas type N	0.05000	0.15000	0.07500
65		maximum growth rate at 0 °C for Flagalletas type P	0.05000	0.15000	0.07500
66		maximum growth rate at 0 °C for DinoFlagellates type E	0.05000	0.15000	0.13200
67		maximum growth rate at 0 °C for DinoFlagellates type N	0.05000	0.15000	0.11300
68		maximum growth rate at 0 °C for DinoFlagellates type P	0.05000	0.15000	0.11200
69		maximum growth rate at 0 °C for Phaeocystis type E	0.05000	0.15000	0.08400
70		maximum growth rate at 0 °C for Phaeocystis type N	0.05000	0.15000	0.07800
71		maximum growth rate at 0 °C for Phaeocystis type P	0.05000	0.15000	0.07800

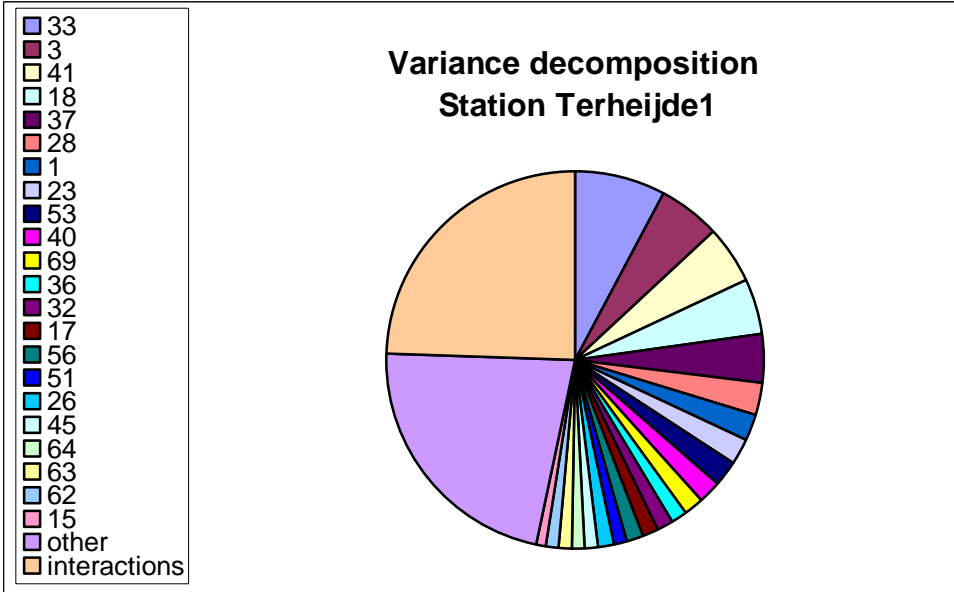
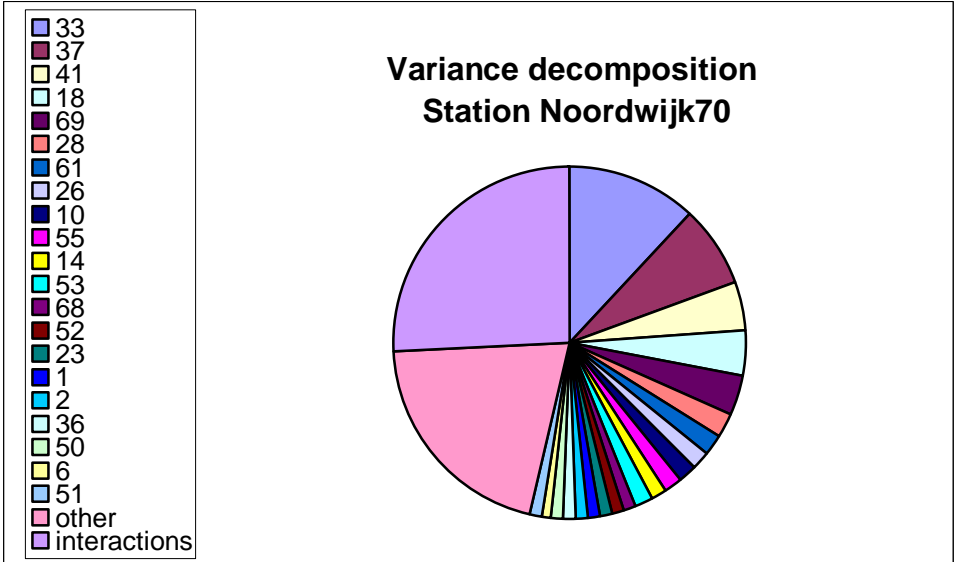
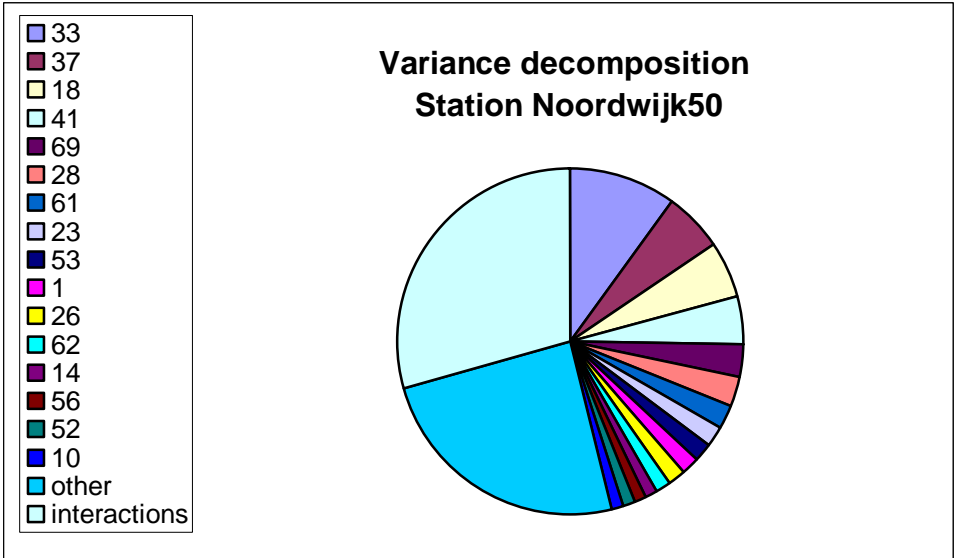
Variance decomposition of the maximal chlorophyll-a concentration:

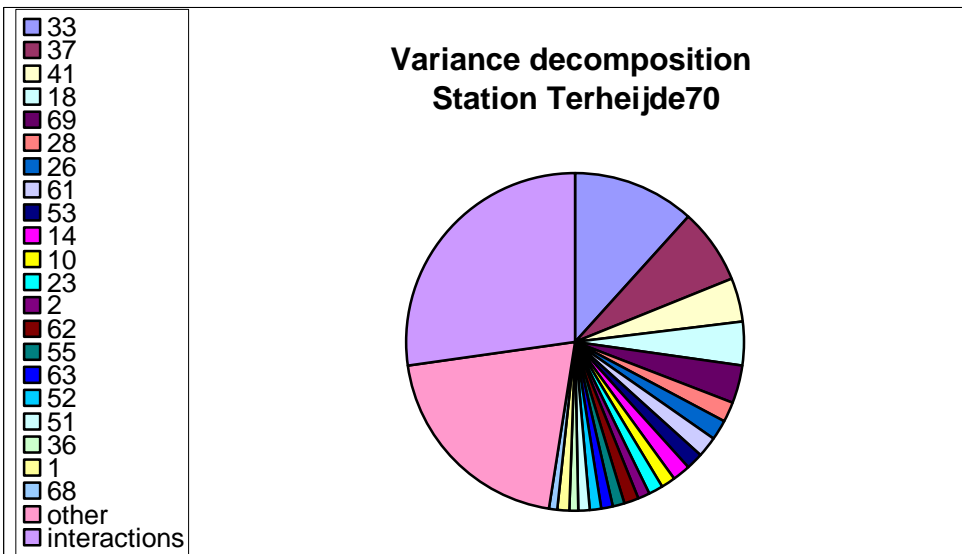
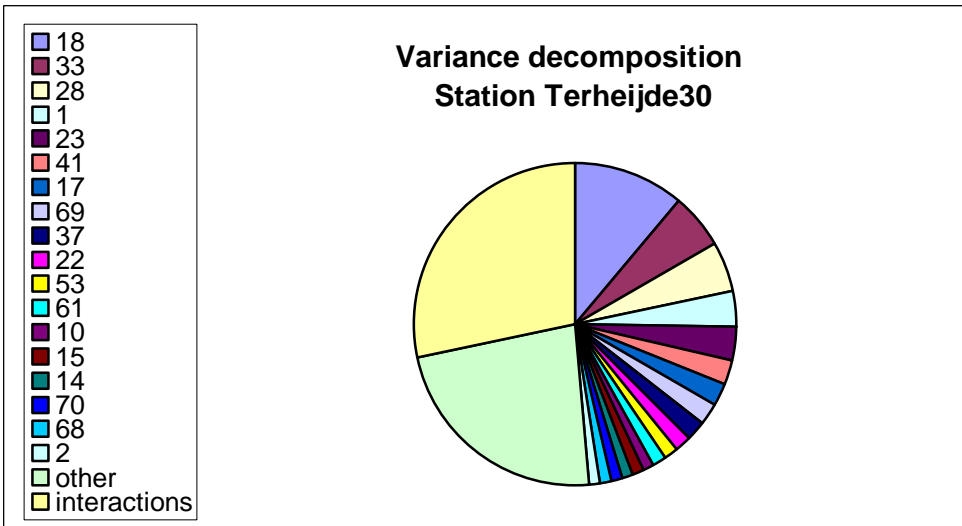
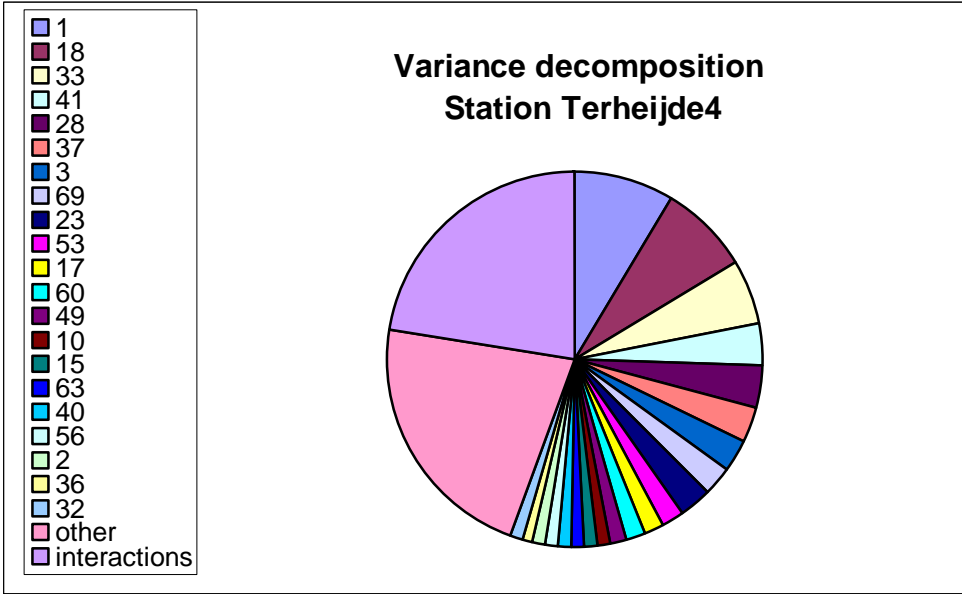


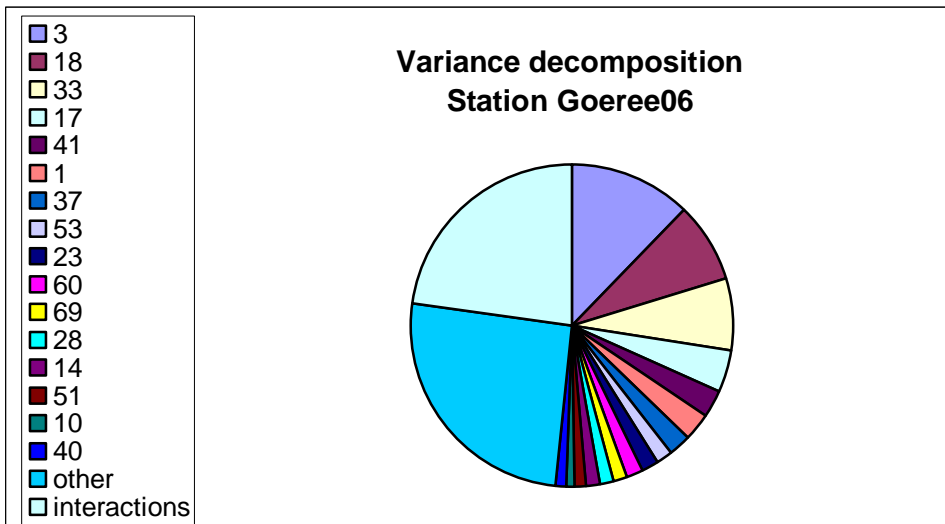












Variance decomposition of the annual average chlorophyll-a concentration:

