

Supporting Information to: Modelling aggregation and sedimentation of nanoparticles in the aquatic environment

A.A. Markus^{a,b,*}, J.R. Parsons^b, E.W.M. Roex^a, P. de Voogt^{b,c}, R.W.P.M. Laane^{a,b}

^a*Deltares, P.O.Box 177, 2600 MH Delft, The Netherlands*

^b*Earth Surface Science, IBED, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

^c*KWR Watercycle Research Institute, Nieuwegein, The Netherlands*

Abstract

This document contains the following supporting information: experimental data used in the modelling, details about the optimisation procedure and error estimates for the model parameters.

1. Laboratory data

This section contains the laboratory data that were used in the calibration of the mathematical model.

Table S1: Laboratory data provided by Quik et al. (2012). In the first series of experiments (in three columns left of the vertical line) water from the Rhine was used, in the second series (right of the line) water from the Meuse was used. The difference between the two water types, as far as the model is concerned, is the concentration of suspended particulate matter (SPM). All concentrations are given in mg/L.

| Initial concentration of nanoparticles SPM | Rhine | | | Meuse | | |
|--|-------|-------|-------|-------|-------|-------|
| | 1.0 | 10.0 | 100.0 | 1.0 | 10.0 | 100.0 |
| SPM | 12.0 | 12.0 | 12.0 | 5.0 | 5.0 | 5.0 |
| Time (d) | | | | | | |
| 0 | 0.61 | 9.79 | 104.3 | 0.5 | 9.36 | 101.7 |
| 1 | 0.40 | 0.94 | 2.17 | 0.24 | 0.36 | 1.14 |
| 4 | 0.23 | 0.117 | 0.136 | 0.155 | 0.066 | 0.093 |
| 7 | 0.126 | 0.059 | 0.029 | 0.10 | 0.026 | 0.083 |
| 10 | 0.087 | 0.042 | 0.022 | 0.098 | 0.021 | 0.040 |
| 12 | 0.078 | 0.040 | 0.026 | 0.096 | 0.020 | 0.032 |

*Corresponding author, telephone +31-88-3358559

Email address: arjen.markus@deltares.nl (A.A. Markus)

Table S2: Laboratory data extracted from Brunelli et al. (2013). The results with North Adriatic Sea water were used, as presented in the supplementary material to the article. The concentration of SPM was assumed to be 0. The data in the table are approximations. The values are given in mg/L.

| | Case 1 | Case 2 | Case 3 | Case 4 |
|--|--------|--------|--------|--------|
| Initial concentration of nanoparticles | 0.01 | 0.1 | 1.0 | 10.0 |
| Time (hours) | | | | |
| 0 | 0.01 | 0.1 | 1.0 | 10.0 |
| 10 | 0.0086 | 0.079 | 0.72 | 4.5 |
| 20 | 0.0074 | 0.069 | 0.55 | 2.0 |
| 30 | 0.0067 | 0.064 | 0.44 | 1.0 |
| 40 | 0.0066 | 0.062 | 0.35 | 0.61 |
| 50 | 0.0059 | 0.052 | 0.25 | 0.30 |

2. Model equations

The main text describes several processes, such as resuspension and mineralisation, that do not play a role in the simulation of the laboratory experiments. This section provides the set of equations that is used in the simulation. The equations are solved numerically using the Euler integration method with a small enough timestep.

The equations are:

$$\frac{dC_{free}}{dt} = -k_{homoagg} \cdot C_{free}^2 - k_{sec_agg} \cdot C_{free} C_{agg} \quad (1)$$

$$-k_{ads} \cdot C_{free} C_{spm} - \frac{w_{s,free}}{H} C_{free} \quad (2)$$

$$\frac{dC_{agg}}{dt} = +k_{homoagg} \cdot C_{free}^2 + k_{sec_agg} \cdot C_{free} C_{agg} - \frac{w_{s,agg}}{H} C_{agg} \quad (3)$$

$$\frac{dC_{ads}}{dt} = +k_{ads} \cdot C_{free} C_{spm} - \frac{w_{s,spm}}{H} C_{ads} \quad (4)$$

$$\frac{dC_{spm}}{dt} = -\frac{w_{s,spm}}{H} C_{spm} \quad (5)$$

The initial conditions were in all cases:

- The initial concentration in the experiment to be simulated was assigned to the concentration of free nanoparticles (C_{free}).
- The initial concentration of the other fractions of nanoparticles (C_{agg} and C_{ads}) was set to 0.
- The initial concentration of suspended particulate matter was set to the concentration indicated in the description of the experiments.

The output of the model includes:

- The concentration of the three distinguished fractions of nanoparticles as a function of time.

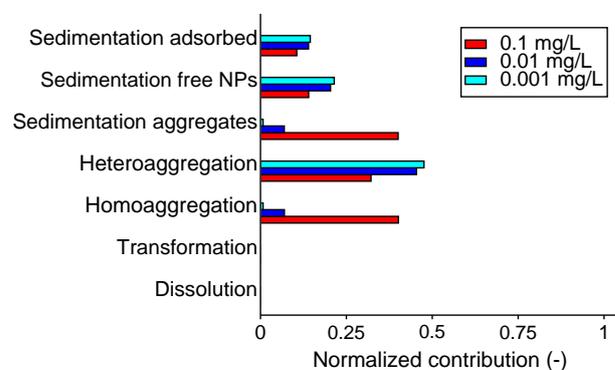


Figure S1: Normalized contributions of the various processes, contributing to the fate of cerium dioxide nanoparticles in Rhine water at initial concentrations in the sub-milligram range.

- The total concentration of nanoparticles still in suspension, that is, the sum of the three fractions. This is a steadily decreasing function of time, as the nanoparticles that settle on the bottom of the cuvette due to sedimentation are no longer in suspension. As there is no resuspension possible (in the context of the simulation of the laboratory experiments), the nanoparticles remain on the bottom.
- The concentration of suspended particulate matter as a function of time.

3. Relative contributions of the processes for low concentrations

To illustrate the importance of the various processes for concentrations that are more likely to occur in the aquatic environment, that is, concentrations in the range of $\mu\text{g/L}$, the calculations were repeated for initial concentrations 0.1, 0.01 and 0.001 mg/L (see Fig. S1). At these low levels the heteroaggregation process quickly becomes much more important than the homoaggregation process. This has consequences for the distribution of the nanoparticles over the three fractions – free, aggregated and adsorbed – so that the sedimentation of the free and adsorbed nanoparticles becomes more important than the sedimentation of aggregated nanoparticles.

4. Automatic parameter estimation

With the given model equations and the experimental data it should be possible to determine values for the process parameters that are in some sense optimal, that is, the solution to the equations approximates the experimental data as close as possible. It should also be possible to apply these values to other experimental data sets, as long as the same nanomaterials are used, if the model equations are indeed an adequate description of the processes. To this end an optimisation procedure was used, but the results required some careful interpretation.

Table S3: Constructed data for the test case, compared with the predicted values, found using the "best" cost function. The constructed data roughly follow an exponential function.

| Time | Constructed data | Predicted value | Relative deviation (%) |
|------|------------------|-----------------|------------------------|
| 0.0 | 1.0 | 1.0968 | 9.7 |
| 1.0 | 0.4 | 0.329 | -17.8 |
| 4.0 | 0.02 | 0.0218 | 9.0 |
| 10.0 | 0.015 | 0.0141 | -6.3 |

4.1. Optimisation procedure

The algorithm used is a constrained minimisation procedure by M. Powell that searches for a solution within given bounds without the need to specify derivatives (Powell, 2009). The latter aspect is important: since the function that is to be minimised is not given explicitly, it is extremely difficult if at all possible, to calculate the derivative directly. Indeed, to evaluate the discrepancy between the modelled total concentration of nanoparticles in the water and the measured concentration, the model equations must be solved first, using a specific set of parameter values.

While the algorithm is capable of finding the minimum for arbitrary continuous functions, the function to be minimised must be an adequate measure for the discrepancy. Several candidates were examined, all based on the idea that the solution to the differential equations should match the measured data as closely as possible. One obvious candidate is the sum of squared errors:

$$f(\underline{p}) = \sum_{i=1}^n (C_{\text{calculated}}(t_i, \underline{p}) - C_{\text{measured}}(t_i))^2 \quad (6)$$

where $C_{\text{measured}}(t_i)$ is the concentration measured at time t_i and $C_{\text{calculated}}(t_i, \underline{p})$ is the concentration calculated via the model equations for the same time, using the vector \underline{p} of parameter values.

The problem with this so-called cost function is that high concentration values are emphasised over low values. For instance, if the measured concentration is 10 mg/L, a relative deviation of 10% gives rise to a contribution of 1 (mg/L)² to the total error, whereas it was 1 mg/L, such a deviation contributes only 0.01 (mg/L)². Since in the available experiments the concentrations vary over several orders of magnitude, this means that only the highest concentrations have appreciable influence on the cost function and those are not in the range that is expected to occur in the environment.

Alternatives included: the logarithm and the square root of the concentrations as well as the absolute value of the differences instead of the square.

The function that was found to give uniform results over the whole concentration range, not emphasizing either high or low concentrations, was:

$$f(\underline{p}) = \sum_{i=1}^n \frac{(C_{\text{calculated}}(t_i, \underline{p}) - C_{\text{measured}}(t_i))^2}{C_{\text{measured}}(t_i)^2} \quad (7)$$

These cost functions were applied to a simple minimisation problem (see Table S3): Find coefficients A, B and C, such that the function $f(t) = Ae^{-Bt} + C$ matches the constructed data as close as possible.

The deviation per data point was examined. The chosen cost function, which measures the relative deviation rather than an absolute deviation, gave fairly uniform relative differences between the model predictions and the constructed data, whereas the other cost functions gave both very small and very large differences. In a numerical experiment with constructed data which significantly deviate from an exponential function, all cost functions showed large relative differences.

4.2. Interpretation of the fitted parameters

The optimisation process is in itself straightforward:

- For each parameter a range is given within which the optimal value is sought.
- For each set of parameters the model equations are solved for all selected cases and the cost function is determined based on the solution and the differences with all measurements. For the experiments with cerium dioxide this means the set of equations is solved using the three initial concentrations for the Rhine and the three initial concentrations for the Meuse. The function describing the difference between the model result and the laboratory experiments is therefore based on 30 data points. It is used to find optimal values for six parameters.

For the titanium dioxide experiments 20 data points are involved and four initial concentrations. Since the natural water that was used in these experiments by Brunelli et al. (2013) was filtered, it was assumed in this work that the processes involving suspended solids should be excluded, resulting in only three parameters to be fitted.

- The algorithm searches for the optimal set of parameter values without further guidance.

The problem is, however, that using the data from the experiments by Quik et al. (2012) the optimal parameter values were found at the boundary of the given range. Widening the range would still result in an apparent optimal value for such a parameter at the boundary, so that further widening seemed necessary. There are several explanations possible for this phenomenon.

First the cost function may have multiple local optima. This is an inherent problem with optimisation algorithms. While an actual proof of the existence of such local minima is difficult to assess as there is no closed form available, we can make their existence plausible in the following way:

- If in the model the sedimentation velocity of suspended solids is sufficiently large, then most if not all suspended solids will have settled before the time of the first data point. For instance, with a sedimentation velocity of 1 m/d the suspended solids in the 3 cm high bottles have a residence time of about 0.03 days. After two hours only a quarter of the initial amount of suspended solids would be left and after one day a minute fraction ($e^{-1/0.03} \approx 3 \cdot 10^{-15}$).
- Under these circumstances the adsorption to suspended solids would be of interest in the simulation only at the start. The sedimentation of aggregates, however, has the same effect on the total concentration, namely a decrease. That means

that the sedimentation of nanoparticle aggregates can compensate for the lack of a substantial contribution by adsorption to suspended solids and subsequent sedimentation of these adsorbed nanoparticles.

So, if the algorithm examines values for the sedimentation velocity of suspended solids that would lead to the quick disappearance of suspended solids, it will end up in the "wrong" region of parameter values, that is: a local minimum.

A second explanation might be that the measurement data do not capture the information about fast processes occurring in the actual laboratory experiments. For instance: the secondary homoaggregation could be a very fast process that occurs on a timescale that is shorter than a day. If so, the data would not adequately reflect the kinetics of this process and any high value could be just as good. To put it in a different way: the solution of the model equations at the times for which there are observed values hardly depends on the value of the parameters that govern this process. Then an optimisation procedure cannot determine any reasonable value for them either.

A third explanation may apply here in combination with the first: within the optimisation algorithm the parameter space is searched in a "linear" way, that is, if the interval within which the optimal values are sought is too wide by an order of magnitude, the algorithm may miss very small values and get stuck in the wrong part of the parameter space. In this case, the interval for the sedimentation velocity of suspended particulate matter was originally set to 0.0 to 10.0 m/d, leading to an unrealistically large value, whereas by setting the interval to 0.0 to 0.01 m/d the final and more conforming value of 0.003 m/d was found. In the numerical simulation the very large sedimentation velocity led to the quick disappearance of suspended particulate matter. The sedimentation of the homoaggregates of nanoparticles then became the only process removing mass, but this proceeds in a more or less similar way. Hence the results for the total mass in the water still looked reasonable.

In short, it is necessary to check the plausibility of the outcome of such optimisation procedures, if at all possible.

4.3. Error estimation

To gain insight in the robustness of the parameter estimation, a bootstrap method was used (Anonymous, 2014). The basic idea of bootstrapping is to construct alternative data sets based on the original data set and the results of the fitting procedure. These alternative data sets are then subjected to the same fitting procedure and will yield alternative sets of parameters. By examining these parameter sets one gets an impression of the spread of the parameters.

Here the following procedure was used:

- First the residuals were determined for the original data set, that is the differences between the best fit and the original data.
- Then alternative data sets were constructed by replacing the original value by the original value plus or minus the residual, with equal probability. In this way 1000 alternative data sets were constructed.
- For each of these alternative sets the parameters were fitted and recorded. From this collection several percentiles were determined: 5%, 25%, 50%, 75% and 95%.

Table S4: Distribution of the model parameters as determined with the bootstrap method for the cerium dioxide experiments. The percentiles are based on 1000 alternative datasets.

| Parameter | Best value | 5% | 25% | 50% | 75% | 95% |
|------------------------------|------------|--------|--------|--------|--------|--------|
| Homoaggregation L/mg.d | 1.77 | 1.34 | 1.67 | 1.89 | 2.07 | 2.34 |
| Secondary aggregation L/mg.d | 2.28 | 1.91 | 2.10 | 2.45 | 2.90 | 3.29 |
| Aggregation to SPM L/mg.d | 0.0073 | 0.0032 | 0.0041 | 0.0055 | 0.0085 | 0.0131 |
| Sedimentation aggregates m/d | 0.24 | 0.22 | 0.23 | 0.24 | 0.32 | 0.32 |
| Sedimentation SPM m/d | 0.003 | 0.0000 | 0.0007 | 0.0018 | 0.0034 | 0.0070 |
| Sedimentation free NPs m/d | 0.002 | 0.0000 | 0.000 | 0.0012 | 0.0034 | 0.0066 |

Table S5: Distribution of the model parameters as determined with the bootstrap method for the titanium dioxide experiments. The percentiles are based on 1000 alternative datasets. Note: the concentration of suspended matter is zero for these experiments, so that the associated process parameters were excluded.

| Parameter | Best value | 5% | 25% | 50% | 75% | 95% |
|------------------------------|------------|-------|-------|-------|-------|-------|
| Homoaggregation L/mg.d | 0.618 | 0.34 | 0.51 | 0.68 | 0.87 | 1.08 |
| Secondary aggregation L/mg.d | 2.254 | 0.085 | 0.67 | 2.22 | 2.64 | 3.39 |
| Sedimentation aggregates m/d | 0.111 | 0.109 | 0.113 | 0.117 | 0.124 | 0.204 |
| Sedimentation free NPs m/d | 0.034 | 0.031 | 0.033 | 0.035 | 0.036 | 0.038 |

The results are shown in Tables S4 and S5. For comparison the value found for the original data set has been added (column "best value").

While some parameters for cerium dioxide display a relatively wide spread, the most influential parameters, the coefficients determining the homoaggregation, have a fairly narrow range, in the order of $\pm 25\%$ for the 25% quantiles.

The ranges for the titanium dioxide experiments are fairly narrow too, especially the parameters controlling sedimentation, but the ranges for the secondary aggregation are an exception. In fact, when the sedimentation velocity of free nanoparticles is set to zero, the optimisation algorithm finds that zero secondary aggregation is the best solution. It is tempting to speculate about a physical meaning of these observations, but there is a risk of over-interpretation. More data are needed for that to be useful.

5. Examination of the experiments with titanium dioxide

As suggested in the main text, there may have been a certain degree of initial aggregation of the nanoparticles in the aggregation and sedimentation experiments by Brunelli et al. (2013), so that the experiments started with larger particles than the nominal size. As a consequence the sedimentation velocity of the "free" nanoparticles that was found by fitting the model was much larger than expected from Stokes' theory.

To examine the influence of the sedimentation of free nanoparticles, the model was also calibrated with this process turned off (see Fig. S3). The results show a worse fit

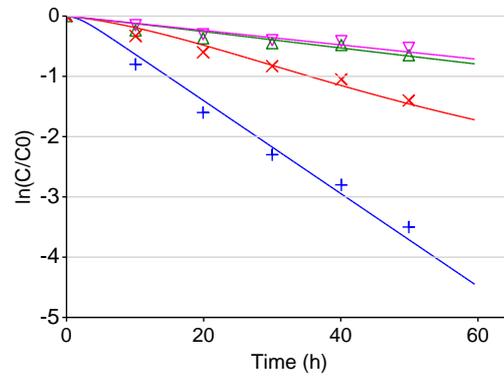


Figure S2: Natural logarithm of the concentration of titanium dioxide nanoparticles as simulated, divided by the initial concentration, using the "best" parameter set for this experiment (see Table S6; adopted from the main text). +: initial concentration 10 mg/L, x: 1 mg/L, Δ : 0.1 mg/L, ∇ : 0.01 mg/L.

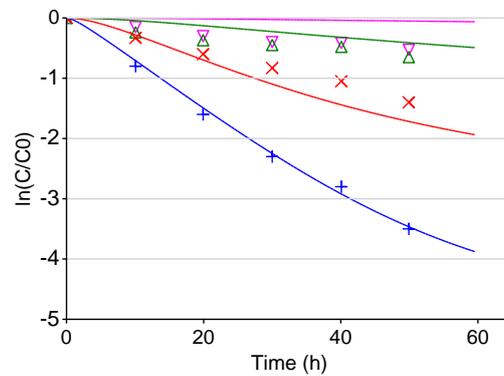


Figure S3: Natural logarithm of the concentration of titanium dioxide nanoparticles as simulated, divided by the initial concentration, using the alternative parameter set for this experiment (see Table S6), that is the sedimentation of free nanoparticles has been eliminated. +: initial concentration 10 mg/L, x: 1 mg/L, Δ : 0.1 mg/L, ∇ : 0.01 mg/L.

Table S6: Fitted parameters for the titanium dioxide experiments by Brunelli et al. (2013) with and without allowing for sedimentation of free nanoparticles.

| Parameter | With sedimentation | Without sedimentation |
|---|--------------------|-----------------------|
| Homoaggregation $k_{homoagg}$; L/mg.d | 0.618 | 3.194 |
| Secondary aggregation k_{sec_agg} ; L/mg.d | 2.254 | 0.000 |
| Settling aggregates w_{sed_agg} ; m/d | 0.111 | 0.134 |
| Settling free NPs w_{sed_free} ; m/d | 0.034 | 0.000 |

Only parameters that were actually fitted are shown. The height of the cuvette was assumed to be 3 cm.

compared to the results shown in Fig. S2, especially for the experiment with the lowest initial concentration. As for the process parameters themselves, if no sedimentation of free nanoparticles is allowed, the rate coefficient for homoaggregation is almost three times as large (Table S6). The fitted sedimentation velocity for the aggregates is slightly smaller than if free nanoparticles are allowed to settle.

While the fitting procedure alone cannot prove that the nanoparticles were already aggregated at the start and therefore could settle faster, the results indicate that including this or a similar process provides a better explanation of the experiments.

6. Availability of the developed programs

The programs developed for the simulations and subsequent analyses are available upon request. There is, however, no user documentation as they were intended for this particular project only.

References

- Anonymous, 2014. Bootstrapping (statistics).
 URL http://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29
- Brunelli, A., Pojana, G., Callegaro, S., Marcomini, A., 2013. Agglomeration and sedimentation of titanium dioxide nanoparticles (n-TiO₂) in synthetic and real waters. *Journal of Nanoparticle Research* 15, 1–10.
- Powell, M., August 2009. The BOBYQA algorithm for bound constrained optimization without derivatives. Internet, dAMTP 2009/NA06.
 URL http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf
- Quik, J. T., Cohen Stuart, M., Wouterse, M., Peijnenburg, W., Hendriks, A. J., van de Meent, D., 2012. Natural colloids are the dominant factor in the sedimentation of nanoparticles. *Environmental Toxicology and Chemistry* 31, 1019–1022.