

The program “NOBLEGAS”

The ultimate tool for the interpretation of noble gas concentrations in water

W. Aeschbach-Hertig, F. Peeters, U. Beyerle, R. Kipfer

Swiss Federal Institute of Technology (ETH) and Swiss Federal Institute for
Environmental Science and Technology (EAWAG), CH-8600 Dübendorf,
Switzerland

The ideas behind this program have been described in:

Aeschbach-Hertig, W., F. Peeters, U. Beyerle, R. Kipfer, 1999. Interpretation of dissolved
atmospheric noble gases in natural waters, *Water Resour. Res.* 35, 2779-2792.
(inverse fitting technique)

and

Aeschbach-Hertig, W., F. Peeters, U. Beyerle, R. Kipfer, 2000. Palaeotemperature
reconstruction from noble gases in groundwater taking into account equilibration with
entrapped air, *Nature* 405: 1040-1044.
(new model for excess air)

Please cite at least one of these papers when using this program to process your data.

1. Introduction

Noble gases in natural waters originate from three major sources: i) dissolution of atmospheric air according to solubility equilibrium, ii) injection of "excess air", and iii) addition of certain isotopes from radiodecay in the water or the aquifer matrix. The first component depends on temperature, salinity, and atmospheric pressure during equilibration of the water. It has been used to calculate the temperature during infiltration of groundwater (noble gas temperature, NGT). The third component is of major importance for He (^3He from ^3H and ^4He from U and Th) and sometimes also affects ^{40}Ar (from ^{40}K). It has extensively been used for dating purposes. Relatively little is known about the origin, composition, and systematics of the second component, "excess air", which is of particular importance for groundwaters, but also occurs in the ocean. It can have the elemental composition of atmospheric air, but in some cases a systematic fractionation of this component was observed, and two models describing this fractionation have been proposed (see chapter 2).

The main idea of the program "NOBLEGAS" is to interpret measured concentrations of He, Ne, Ar, Kr, and Xe in terms of the model parameters temperature (T), salinity (S), pressure (P), excess air (A), and fractionation (F). This interpretation is based on a non-linear, error weighted least squares inversion of the models. Measured concentrations and their errors are entered via a text file (see chapter 3). The actual least square fitting is done by the MATLAB program "Noble-gas", which has a user-friendly interface, where the user can select the model and any combination of fit parameters (chapter 4). The program produces a large outputfile, which in addition to the best parameter values and their errors contains the residuals (deviation between measured and modeled concentrations), the χ^2 -values (weighted sum of the residuals), the respective probability from the χ^2 -distribution, the correlations between the parameters, the non-atmospheric contribution to He, and many other data (chapter 5).

2. Theory

Noble gas concentrations in surface waters are usually close to equilibrium with the atmosphere. The respective equilibrium concentrations can be calculated from the solubilities given in several literature sources (the most practical source are the papers by R. Weiss: *Deep-Sea Res.*, 1970, 17: 721-735; *J. Chem. Eng. Data*, 1971, 16: 235-241; 1978, 23: 69-72). In groundwaters, one almost always finds a substantial excess above these equilibrium concentrations, the "excess air" component (Heaton and Vogel, *J. Hydrol.*, 1981, 50: 201-216). For a long time it was assumed that this component is pure atmospheric air due to complete dissolution of trapped air bubbles in the ground. However, newer noble gas studies showed that this assumption is not generally valid. More complete models, including fractionation of the excess air component, are thus needed to describe atmospheric noble gases in groundwaters.

The first general model was proposed by Stute et al., 1995 (Science, 269, 379-383). It explains the fractionation of the excess air by diffusion controlled partial re-equilibration (PR model) of the initial amount of excess air. It may be written as:

$$C_i = C_i^*(T, S, P) + A z_i e^{-F \frac{D_i}{D_{Ne}}} \quad (i = \text{He, Ne, Ar, Kr, Xe}) \quad (1)$$

where $C_i^*(T, S, P)$ are the moist air solubility equilibrium concentrations in $\text{cm}^3 \text{STP/g}$ as functions of water temperature, salinity, and total atmospheric pressure, A is the STP-volume of dry "excess air" per g of water, z_i is the volume fraction of gas i in dry air, F is the fractionation parameter describing the degree of diffusive re-equilibration, and D_i are the molecular diffusion coefficients.

An alternative model has been proposed by Aeschbach-Hertig et al. (submitted), which explains the fractionation as the result of partial dissolution of entrapped bubbles of air and subsequent equilibration between water and remaining entrapped gas bubbles in a closed system (CE model). The respective model equation is:

$$C_i(T, S, P, A, F) = C_i^*(T, S, P) + \frac{(1-F)Az_i}{1 + FAz_i/C_i^*} = C_i^*(T, S, P) \frac{1 + Az_i/C_i^*}{1 + FAz_i/C_i^*} \quad (2)$$

For the discussion of the physical meaning of the fractionation parameter F in this model, the following definitions are important:

The fractionation factor F is given by $F = v/q$, where $v = V_g^0/V_g$ is the ratio of the entrapped gas volume after partial dissolution and equilibration to the initially entrapped air volume, and $q = (P_g - e)/(P - e)$ is the ratio of the dry gas pressure ($e =$ saturation water vapor pressure) in the trapped gas to that in the free atmosphere.

In the program, a third model is implemented, which is intended to describe degassing as it may occur during sampling, when noble gases are stripped into bubbles formed e.g. by CO_2 . In this case, which has been described by Stute, 1989 (PhD thesis, Univ. Heidelberg), it appears reasonable to assume a diffusive gas loss into an initially noble gas free gas phase. Hence, in contrast to the re-equilibration model (eq. 1) not only the excess air component but the total noble gas concentrations are affected. The model equation may be written as:

$$C_i = \left(C_i^*(T, S, P) + A z_i \right) e^{-F \frac{D_i}{D_{Ne}}} \quad (3)$$

It should be noted, that this model does not describe an effect related to the formation of excess air, but simply tries to account for an artefact produced during sampling. It should be used with caution and in our experience rarely offers a good explanation for "strange" samples.

In principle, equations (1) to (3) are non-linear equation systems for the five observables He, Ne, Ar, Kr, and Xe with the five unknown parameters T , S , P , A , and F . However, because the measured He concentration is usually affected by non-atmospheric sources, it cannot be described by these models and the number of applicable equations reduces to four. In the practice of the noble gas paleothermometer, the parameters P and S are usually well known ($S = 0$ for meteoric recharge and P is given by the altitude of the recharge area), and only T , A , and possibly F are unknown (F

$= 0$ reduces to unfractionated excess air). Thus, the equation systems are over-determined and can be solved by least squares techniques.

The goodness of fit that is achieved can quantitatively be judged based on the χ^2 -test. χ^2 is the sum of the squared deviations between the modeled and measured concentrations, weighted with the experimental 1- σ -errors. The expected minimum value of χ^2 is the number of degrees of freedom ν , which is the difference between the number of applied constraints (measured concentrations) n and the number of free parameters m . Because ν is usually very small for a single sample, it is useful to enhance the significance of the χ^2 -test by applying it to entire sample sets. Fitting the same model to N samples of a data set is equivalent to fitting a model with $m \cdot N$ free parameters to $n \cdot N$ data points. The χ^2 -value for the whole data set is the sum of the χ^2 -values of the individual samples, and the number of degrees of freedom is $\nu \cdot N$. The probability p for χ^2 to be higher than a given value can be obtained from the χ^2 -distribution. Models which yield too low probabilities should be rejected. A reasonable criterion might be $p < 0.01$, although in Aeschbach-Hertig et al. (1999) we rejected models with probability $p < 0.05$, which is a rather strong criterion.

The best way to estimate the uncertainties of the derived model parameters, in particular the noble gas temperature (NGT), is a numerical propagation of the experimental uncertainties of the concentrations through the inverse procedure by Monte Carlo simulations. This can be done by generating a number of artificial sets of concentrations from each measured sample, randomly distributed around the measured values according to the experimental errors. For each generated data set, the optimal parameter values are calculated, and their distribution defines the uncertainty. However, if the experimental uncertainties are independent and normally distributed, the uncertainties of the estimated parameters can much easier be derived from the covariance matrix calculated during the least squares fitting. In our experience, the two methods usually yield almost identical results. Hence, the initially included option for Monte Carlo simulations has been dropped in the current version of the program "Noblegas". However, in rare cases fitting equation (2) with T , A , and F as free parameters leads to a strong correlation between T and A and unrealistically large errors from the covariance matrix. In these cases, only Monte Carlo simulations yield reasonable error estimates. For this reason, it is planned to re-implement the Monte Carlo option in future versions of the program.

The errors obtained from the covariance matrix or from Monte Carlo simulations do not take into account how well the model actually fits the data. Thus, standard least square algorithms often scale the errors of the parameters with the factor $(\chi^2/\nu)^{1/2}$. This scaling implicitly assumes that the model is correct but the stated experimental errors were too small and should better be estimated from the deviation of the data from the model. If you have confidence in the experimental errors, you may prefer to use χ^2 only to test the applicability of the model, but not for scaling of the derived errors. Relatively large, but still acceptable, χ^2 -values for some individual samples from a population which as a whole is well described by a certain model may simply be the result of statistical fluctuations. Samples with larger χ^2 -values ($p \ll 0.01$) may rather not be interpreted at

all (due to lack of an appropriate model) rather than stating a noble gas temperature with a very large uncertainty.

While He concentrations can usually not be used for the inversion of the models, the model parameters derived from the concentrations of the heavier noble gases yield the best estimate for the atmospheric (equilibrium plus excess air) component of the measured He. Thus, the difference between measured and modeled He concentration is the non-atmospheric component. The program “NobleGas” invites the input of the measured concentrations of He, even if they are not used for the inversion, and calculates the best estimates of the respective non-atmospheric components.

3. Input

The input is made via a tab-delimited text file which contains all measured noble gas concentrations plus initial/default values for the parameters. The file has 17 columns and the first row is used for a header.

Table 1. The input file

column	explanation
Sample	Sample name. Text.
Nr	Identification number of sample. Must be an integer number.
He	Measured He concentration in ccSTP/g. Often not used as a constraint in the fitting procedure, but to calculate the non-atmospheric component.
err He	Experimental error of the He concentration in ccSTP/g.
Ne	Measured Ne concentration in ccSTP/g.
err Ne	Experimental error of the Ne concentration in ccSTP/g.
Ar	Measured Ar concentration in ccSTP/g.
err Ar	Experimental error of the Ar concentration in ccSTP/g.
Kr	Measured Kr concentration in ccSTP/g.
err Kr	Experimental error of the Kr concentration in ccSTP/g.
Xe	Measured Xe concentration in ccSTP/g.
err Xe	Experimental error of the Xe concentration in ccSTP/g.
T	Initial and default value for the parameter temperature in °C.
S	Initial and default value for the parameter salinity in g/kg.
P	Initial and default value for the parameter pressure in atm (usually from altitude).
A	Initial and default value for the parameter excess air in ccSTP/g (usually 0).
F	Initial and default value for the parameter fractionation (usually 0).

Missing input data should be replaced by 0 in the input file. They are then not used in the calculations. It is possible to mix samples with different numbers of measured noble gases in one file.

4. The MATLAB program “NOBLEGAS”

The actual least square fitting is done by the MATLAB program “NOBLEGAS”. This program is the heart of the whole method. It has a user-friendly interface, where the user can select several options, e.g. the measured noble gases to fit or the combination of fit parameters.

The program is started by typing "noblegas" in MATLAB command window. Make sure that the folder where the program is located is included in the MATLAB's Path list, otherwise it will not be found. A new window is opened, from which the processing of noble gas data sets is controlled. The following options can be chosen:

- Data File: Input data file according to section 3.
- Solubilities: Choose the solubility data base (see Aeschbach-Hertig et al., 1999). Weiss is default.
- Tracers: Check the gases that shall be fitted. Default is Ne, Ar, Kr, Xe.
- Functions and Parameters: In the window below, you can define the combinations of free parameters and the model function that shall be fitted. The three currently available model equations are (see chapter 2):
 - (1) Closed system **E**quilibrium (CE or "Aeschbach" model, see equation (2) above).
 - (2) **P**artial **R**e-equilibrium (PR or "Stute" model, see equation (1) above).
 - (1) **P**artial **D**egassing (PD, for degassed samples or "Stute 2" model, see eq. (3)).

The required input in this field is a list of specific "models" that shall be fitted to the data, e.g., “taf-1” means that temperature, excess air, and fractionation are free parameters and that the CE model equation is used. “af-2” would mean that temperature is taken from the input file, and only excess air and fractionation are free parameters in the PR model equation. Note that if the fractionation parameter F is not fitted, and set to zero in the input file, the choice of the model equation is arbitrary, as they all become identical. If no number is given (e.g., simply “ta”), model 1 is used. There is a default list of often used models.

Adding "u" (e.g. "tafu-1") gives an unconstrained fit, otherwise the parameters are constraint to physically realistic values (positive). If this option is used, the results for the parameters should be checked for "impossible" results (typically negative A or F values).

- Secant or simplex fitting: Choose the search algorithm to find the χ^2 -mimum. Secant is default. The simplex option has not been thoroughly tested. It seems to work better for constrained fits in cases when a parameter runs into a boundary. For most samples, however, it seems that the secant algorithm converges faster.
- Data Format: Pressing this button provides information on the input file format, see section 3.
- Start Fit: Starts the calculation. If the input file has a wrong format, an alert is displayed in the “Error Inf” window. Tip: Empty lines at the end of input files created with "Microsoft Excel" have sometimes caused errors ("wrong number of columns"). Check your input file with a text editor and remove trailing lines if this problem occurs.

During the calculation, messages are written to the MATLAB command window every 5 seconds or so, to provide information on the progress of the fitting. Error messages (e.g., not converged) are written to a log-file on disk.

5. Output

The output is a large tab-delimited text file which contains the best parameter values and their errors, the residuals (deviation between measured and modeled concentrations), the χ^2 -values (weighted sum of the residuals), the respective probability from the χ^2 -distribution, the correlations between the parameters, the non-atmospheric component of He, and many other data in a total of 71 columns (see table 2). The results for each applied model are grouped, with the data of each sample in one row. At the end of each group, a summary for that model is given (total χ^2 and respective probability). Thus the number of data rows in the file is the number of samples +1 times the number of applied models, which can be quite large.

The columns entitled “no of freedom”, “chi2” and “prob” contain the relevant information for the statistical judgement of the fit. In the last row, these statistical results for each sample are summarized to yield a statistical evaluation of the whole data set. Although the data set evaluation yields a very convenient summary parameter to judge different models, the results for the individual samples should also be studied with some care. Sometimes, bad results for the complete data set are due to only a few very bad samples. Usually, such samples cannot be fitted by any model, and they thus yield little information to decide which model describes the data better. Such samples may be the result of bad sampling conditions (e.g., degassing) and may have to be thrown out of the evaluation. Furthermore, attention should be given to the physical reality of the individual fits. Especially, fits with negative results for the fractionation parameter F are questionable, irrespective of the applied model. In such cases, the samples should be fitted with unfractionated excess air only. Negative excess air (A) values indicate degassing, which may be approximated by model 3.

At the beginning of the output file, a general header lists the program version used and the user choices made in the MATLAB program, for later reference. The headers of the 71 output data columns and their meanings are:

Table 2. The output file

column	explanation
function	Model function = choice of free parameters and model equation.
sample_id	Sample name. Text. As in input file.
sample_nr	Identification number of sample. As in input file.
no of freedom	number of degrees of freedom (no. of constraints – no. of parameters).
chi2	Minimum value of χ^2 found by the fitting procedure.
prob [%]	Probability for a larger χ^2 -value from the χ^2 -distribution, in %.
T [°C]	Optimal value for the parameter temperature.
errT (scaled) [°C]	Uncertainty of T, scaled with $(\chi^2/n)^{1/2}$ to account for goodness of fit.
errT (cov) [°C]	Uncertainty of T from the covariance matrix (error propagation).
S [g/kg], errS (scaled), errS (cov)	Optimal value for the parameter salinity and its uncertainty (both scaled and directly from the covariance matrix).
P [atm], errP (scaled), errP (cov)	Optimal value for the parameter pressure and its uncertainty (both scaled and directly from the covariance matrix).
A [ccSTP/g], errA (scaled), errA (cov)	Optimal value for the parameter excess air and its uncertainty (both scaled and directly from the covariance matrix).
F, errF (scaled), errF (cov)	Optimal value for the parameter fractionation and its uncertainty (both scaled and directly from the covariance matrix).
convergence	Convergence of fit (0:no,1:yes)
corrTS, TP, TA, etc.	Correlations between pairs of parameters
excess air [% Ne]	Excess air (A) expressed as Ne excess relative to equilibrium in %.
rad. He (m-p) [ccSTP/g]	Radiogenic He calculated as difference between measured (m) and model predicted (p) He concentration.
He (m-p)/m [%] ... Xe (m-p)/m [%]	Relative deviation between measured (m) and predicted (p) concentrations of the different gases (normalized residuals).
He equ [ccSTP/g] ... Xe equ [ccSTP/g]	Calculated solubility equilibrium concentrations (functions of T, S, P).
He p [ccSTP/g] ... Xe p [ccSTP/g]	Predicted (modeled) concentrations of the atmospheric (equilibrium + excess air) noble gases.
He errp [ccSTP/g] ... Xe errp [ccSTP/g]	Errors (uncertainties) of the predicted concentrations, calculated from the covariance matrix.
He m [ccSTP/g] ... Xe m [ccSTP/g]	Measured concentrations.
errHe [ccSTP/g] ... errXe [ccSTP/g]	Errors (experimental uncertainties) of the measured concentrations
solubility He ... solubility Xe	Code for the solubility data set that was used: 1 = Weiss; 2 = Clever; 3 = Benson.
q_CE-Model	Value of parameter q in CE fractionation model (see after eq. 2 above)
v_CE-Model	Value of parameter p in CE fractionation model (see after eq. 2 above)