Navy Position on Tarawa Terrace Water Modeling
April, 2009

General Comments:

ATSDR's "Response to the Department of the Navy's Letter on Assessment of ATSDR Water Modeling for Tarawa Terrace (March, 2009)" gives a detailed response to Navy concerns expressed by letter on June 19, 2008. The Navy concurs that ATSDR has followed accepted procedures for contaminant transport modeling, and the approach is in line with similar peer-reviewed modeling efforts. Our concerns have focused on the interpretation of the model results, and the need to convey an estimate of model accuracy to the public, epidemiologists, and other users of the model. This letter serves to document the Navy's current position in light of the ATSDR response. The Navy feels that continued discussion of these issues will be beneficial, as lessons learned from Tarawa Terrace can be applied to Hadnot Point. A brief summary is given below, followed by more detailed comments.

- Relative exposure. Navy understands that the epidemiological study will emphasize the relative level of exposure to PCE rather than the actual (absolute) level of exposure. The accuracy of the model is still a concern because measured concentrations can be expected to fluctuate over time, without producing a consistent relative difference between measured and model-derived PCE concentrations. In addition, the reported range in model output (Figure I29 of the Tarawa Terrace water modeling report and attached Figure 2) allows for significant uncertainty in ranking individuals according to their level of exposure.
- Presenting uncertainty to stakeholders. Navy is pleased to see that the web site has been revised from showing single value estimates of PCE concentrations to showing ranges of concentrations. For greater understanding by the public, we believe a linear scale would be more effective than the logarithmic scale currently shown (Figure I29 of the Tarawa Terrace water modeling report). In addition, we suggest using an estimate of model accuracy, rather than precision, as the basis for establishing the range of potential PCE exposures. Under this approach (based on comparing model-derived concentrations to measured concentrations), the PCE concentration range would vary by a factor somewhat greater than 10, rather than a factor of 2 or 2.5 as is currently shown in Figure I29. A graph depicting the range in PCE concentrations based on estimated model accuracy would help model users understand the output and put the model results in perspective.
- Lack of measured data. Navy and ATSDR agree that there was not enough measured PCE data for a meaningful verification step. Since measured PCE concentrations are only available in the 1980's, model output from the early 1980's back to the 1950's cannot be compared to actual PCE data (see attached Figure 1). Confidence in the accuracy of the model is lessened due to the lack of measured PCE data and the length of time (roughly 30 years) over which model output cannot be verified.

The words "accuracy" and "precision" are used in this paper according to the following definitions: **Accuracy** is the extent of agreement between model output and measured data.

Accuracy is estimated by comparing the model to the real world. **Precision** is the extent of agreement among various model runs. Precision is estimated by comparing one model run to another; for example, during Monte Carlo analysis.

Specific Comments:

Section 5.2. ATSDR recognizes the Navy's concern that the calibrated model was not compared to an independent data set, and ATSDR prefers to use the term "verification" rather than "validation." ATSDR says, "The field data set at Tarawa Terrace used for model calibration was not of sufficient quantity and was too compressed in time to implement a verification procedure." Navy agrees with ATSDR that the data are insufficient to permit verification, and we did not intend to suggest that the existing data should have been split into separate calibration and verification data sets. Navy was merely pointing out, in agreement with ATSDR, that there was not enough data for a meaningful verification step. This missing step is significant because the model was never put at risk of failure. That is, once the parameters were determined through calibration, model output could not be compared to actual data to determine the accuracy of the model from the early 1980's back to the 1950's. Confidence in the accuracy of the model is lessened due to the lack of measured data and the length of time (roughly 30 years) over which model output cannot be verified (see attached Figure 1).

Section 6.2. The Navy did not intend to imply that ATSDR should have chosen a different standard for calibration. Instead our comment was simply noting that the chosen calibration standard meant that a model-derived PCE concentration could be roughly 3 times higher or 3 times lower than the measured concentration and still be within the calibration range. This gives an idea of the expected accuracy of the model.

Section 7.2. Navy concurs that the ATSDR calibration process follows established model calibration procedures and generally matches the approach from other similar peer-reviewed reports. We are well aware of the USGS model for NAS Jacksonville, and because of its limitations, the Navy has used the model as a guide in designing field investigations, such that ultimate risk assessment and cleanup decisions are based on sample data rather than model predictions. For example, the NAS Jacksonville model predicted that chlorinated contaminants were possibly reaching the St. John's River, leading the Navy to sample groundwater discharge to the river. The model was useful in guiding the site investigation, but the predicted concentrations by themselves would have been misleading rather than useful for exposure assessment.

Navy agrees with ATSDR that a model should be sufficiently calibrated to support its intended use. Models are never completely accurate, so it is essential to understand whether or not a model is sufficiently accurate for its intended use. ATSDR says the USGS model for NAS Jacksonville was "…totally acceptable to the DON." While it is true that the model was acceptable for its use in guiding the site investigation, the Navy did not accept the model results for use in risk assessment or to design a remediation system. Navy concurs with ATSDR's conclusion that the Tarawa Terrace model and the NAS Jacksonville model are "…of the same order of accuracy and quality…." However, the key difference is the intended use of each

model. As described above, the NAS Jacksonville model was used to guide field investigations, and model predictions were ultimately verified or refuted by real world sampling and analysis. In contrast, the Tarawa Terrace model results are intended to be used alone to provide monthly PCE exposure concentrations for an epidemiological study.

With respect to the epidemiological study, the Navy understands ATSDR's distinction between absolute and relative levels of exposure. Nevertheless, we are concerned with the ATSDR statement, "…exposed individuals are, in effect ranked by exposure level and maintain their rank order of exposure level regardless of how far off the estimated concentration is to the "true" (measured) PCE concentration." The Navy disagrees that inaccurate model concentrations would necessarily rank individuals accurately according to their relative level of exposure. For example, from the mid-1960's to the late 1970's, the range between the 2.5 and the 97.5 percentile of Monte Carlo simulations (attached Figure 2) is roughly 30 to 90 ug/L PCE. Because there is almost complete overlap in the range of model output, accurate ranking of individuals does not seem possible during this time period.

ATSDR says, "This rank order of exposure level is preserved regardless of whether the mean or the upper or lower 95% of simulated levels are used to estimate the monthly average contaminant levels." This approach of consistently using the mean or the upper or lower 95% concentration would not be a proper interpretation of the model output. Given the uncertainty in model results, the actual concentration could fall anywhere within (or even outside) the calculated range of model concentrations. In reality, actual concentrations can be expected to bounce back and forth over time. This type of scatter is typical of environmental data, and it is certainly observed in the measured PCE concentrations at Tarawa Terrace (see Section 8.2). Such scatter would not create a consistent or predictable difference between model output and actual exposures. Thus, it cannot be assumed that actual exposure concentrations would consistently fall along the mean, the upper or lower 95% concentration, or any other consistently chosen location within the estimated range of model output.

ATSDR says the goal of the epidemiological study is "…to evaluate exposure-response relationships to determine whether the risk for a specific disease increases as the level of the contaminant (either as a categorical variable or continuous variable) increases." This raises concerns because model output (Figure I29) indicates uncertainty in PCE concentration trends over time. This uncertainty applies to long term trends over many years, as well as short term trends over a few months. For example, from the mid-1960's to the late 1970's, the range between the 2.5 and the 97.5 percentile of Monte Carlo simulations (Scenario 2) can accommodate either an overall decreasing or increasing trend in PCE concentration. As another example, for the calibrated versus the mean concentrations (Figure I29), there are times when the calibrated concentration is decreasing while the mean concentration (Scenario 2) is increasing, and vice versa. When the trends do not agree, it is not clear that the appropriate trend can be identified for the epidemiological study.

The effects described above can be seen more readily if concentrations are plotted on a linear scale (see attached Figure 2) rather than a logarithmic scale (Figure I29). Overall, the Navy's main concern is that the model accuracy be estimated and the inherent inaccuracies be taken into account by the epidemiological study. The Navy would welcome further discussion with

ATSDR to reach a better understanding of these issues and promote effective use of the model output.

Section 8.2.
Navy stands behind its assessment of the accuracy of the Tarawa Terrace model. As stated above, we are not saying that the calibration targets should have been different, nor are we making an issue of when the calibration targets were chosen. Nevertheless, we do observe that the calibration targets were chosen to be ± ½ order of magnitude, and much of the output (12% at the WTP and 53% at the water supply wells) did not fall within this range (see attached graphs of simulated versus observed concentrations of PCE in water supply wells RW2, TT-23, TT-25, TT-26, and TT-54, Figures F13 through F17 of the ATSDR modeling report). These results indicate that the model is less accurate than ± ½ order of magnitude, and there are no other measured PCE concentrations to support it being more accurate than this (probabilistic analysis compares model runs against each other, so it is a measure of model precision, not accuracy).

ATSDR says that during January and February 1985 at well TT-26, "the field data varied by as much as 2.5 orders of magnitude." Navy recognizes the variability in the field data, and this kind of variability is expected. In our experience at many hundreds of site across the country, measured concentrations of contaminants in groundwater vary significantly and somewhat unpredictably over time. The Navy acknowledges that the measured PCE concentrations are not completely accurate with respect to the true exposures, but this shortcoming does not make the model output more accurate. Instead, it is just another instance of uncertainty in the attempt to recreate historical exposure concentrations at Tarawa Terrace.

Navy is pleased to see that ATSDR has revised its modeling website to include a range of PCE concentrations. However, we feel it would be more representative and useful for the epidemiological study to express this range based on the accuracy of the model, rather than its precision. When the range is based on the probabilistic analysis, this demonstrates model precision, which depends on the chosen distribution of the input parameters. Since Appendix I was not available for comment in June, 2008, the Navy did not address input parameter distributions at that time. Rather than focus now on the input distributions, the Navy proposes that the range in PCE concentration be estimated and displayed based on model accuracy. This would lead to a range that is based on a factor somewhat greater than 10, rather than a factor of 2 or 2.5 (see Section 16.2 below).

For the water modeling website and other presentations to the public, Navy suggests replacing the logarithmic scale of Figure I29 with a linear scale (see attached Figure 2), which can be understood more clearly by the general public.

Section 10.2.
In commenting on DNAPL, the Navy was addressing the difficulties in estimating DNAPL mass and distribution, which affects mass loading into an aqueous phase model. Navy did not comment on density-affected flow in the aqueous phase, and agrees with ATSDR and others that this effect does not need to be considered.

Section 11.2.
The intent of the Navy's comment was simply to note the magnitude of the difference between observed and simulated groundwater elevations. Navy acknowledges ATSDR's description of the inaccuracies in using pressure gages and airlines to measure water levels. While errors in field measurements and sample results certainly impact model calibration, these errors to do not increase the accuracy of the model. Instead, they are another source of uncertainty.

Section 12.2.
ATSDR says, "ATSDR is in agreement with the Navy that PCE concentrations at the WTP are dependent on the pumping rates assigned to the water-supply wells" and "ATSDR shares the DON's concern that simulated operations may not match historical operations." ATSDR indicates that the uncertainty in pumping rates is addressed by the probabilistic analysis, which includes pumping rate as an uncertain model parameter. Appendix I5 of the Tarawa Terrace modeling report shows two scenarios -- Scenario 1 with pumping uncertainty excluded and Scenario 2 with pumping uncertainty included. If uncertainty over pumping rates is a valid concern, it is unclear why Scenario 1 is provided, especially since Appendix I5 is now posted to the public website.

Section 13.2.
Navy did not intend to suggest that 510 was not enough realizations. However, we continue to be concerned that 330 out of 840 realizations did not produce physically viable results. With this being the case, do the input parameter distributions adequately represent site conditions? This appears to be an additional area of uncertainty.

Section 15.2.
ATSDR says, "…the models (flow, transport, and mixing) are sufficiently calibrated, given the quantity and accuracy of data provided and the intended use of the simulated historically reconstructed concentrations for the epidemiological study…" Navy agrees that ATSDR applied proper methods for calibration. However, it is not clear that the model output is sufficiently accurate to support the epidemiological study (see Section 7.2 above).

Section 16.2.
The probabilistic analysis indicates that the range in PCE concentration varies by a factor of about 2.5 when pumping is considered an uncertain variable. ATSDR says, "This is well within acceptable confidence limits for the intended use of the reconstructed PCE concentrations needed by the epidemiological case-control study." The factor of 2.5 comes from probabilistic analysis in which model runs are compared to each other, so the 2.5 factor reflects model precision. Since the epidemiological study requires monthly exposure concentrations of PCE in drinking water, it is less important to understand how the model compares to itself and more important to estimate how the model compares to the real world. Therefore, a more representative PCE concentration range may be derived from model accuracy, which can be estimated by comparing the model PCE concentrations to measured PCE concentrations. As described in Section 8.2 above, this comparison gives a range of PCE concentrations that would vary by a factor somewhat greater than 10; that is, the upper value in the range would be more than 10 times higher than the lower value.

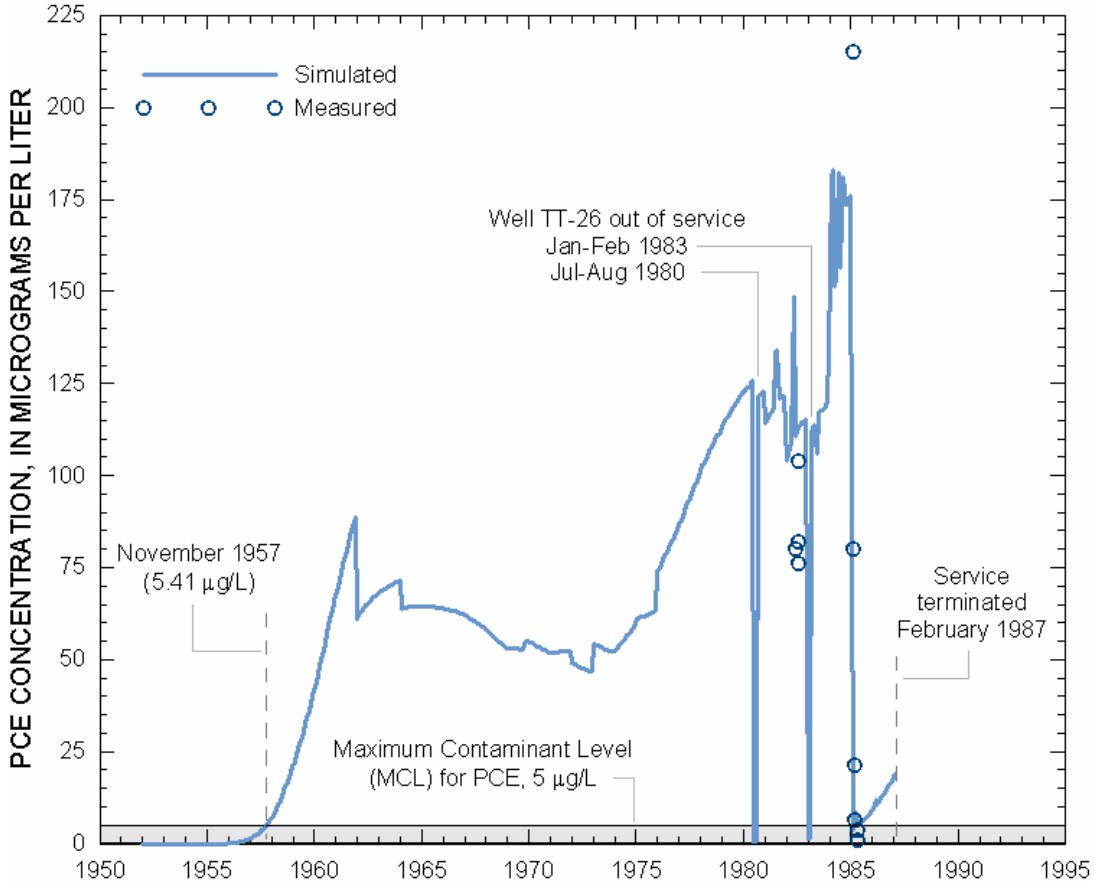

Figure 1. Simulated and measured concentration of tetrachloroethylene (PCE) in finished water at the Tarawa Terrace water treatment plant (from Tarawa Terrace Chapter A report).

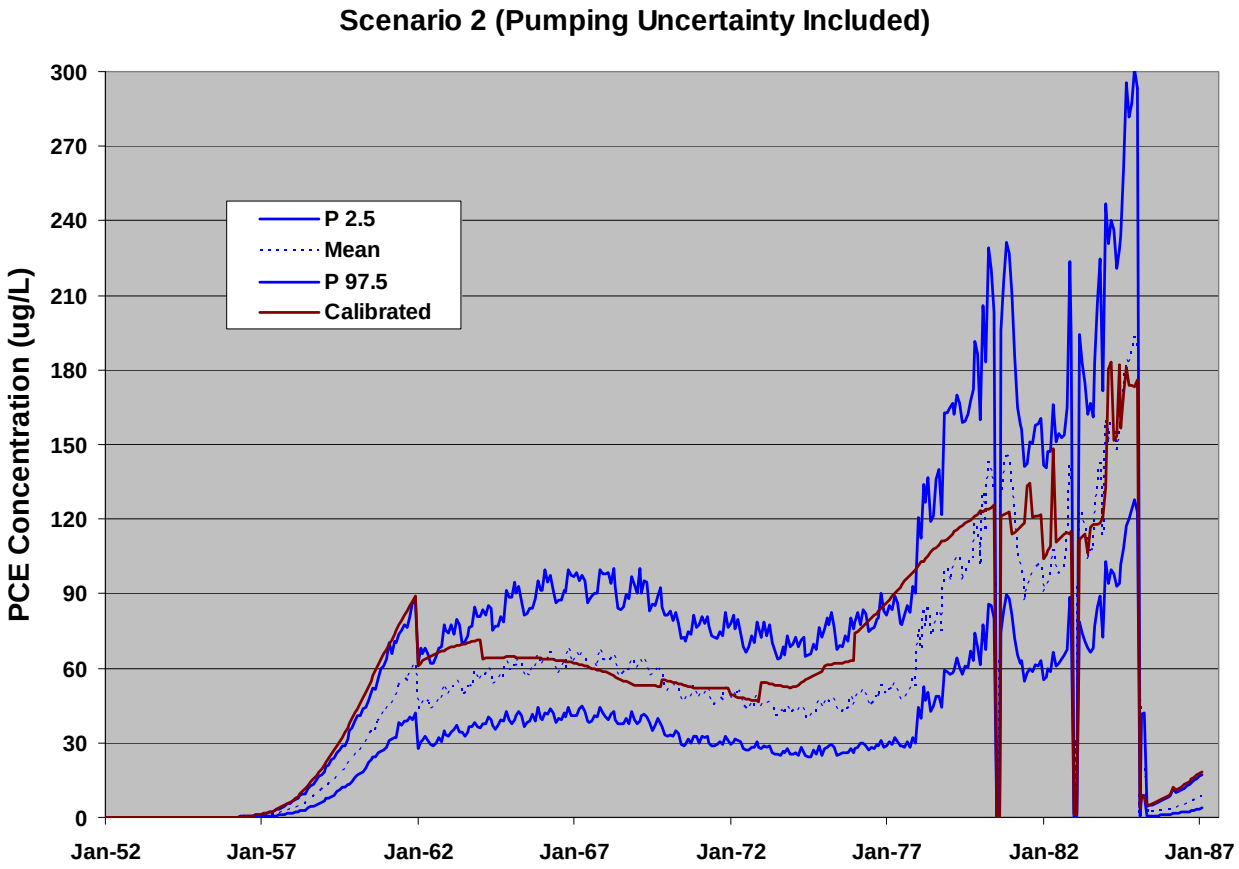

Figure 2. Model-derived concentrations plotted on a linear scale for Scenario 2, in which pumping uncertainty is included (see Chapter I of the ATSDR water modeling report). P 2.5 is the 2.5 percentile of Monte Carlo simulations; P 97.5 is the 97.5 percentile of Monte Carlo simulations; and the mean refers to the mean value of concentration derived from Monte Carlo simulation. The calibrated value is the calibrated concentration using MT3DMS model in a deterministic analysis.

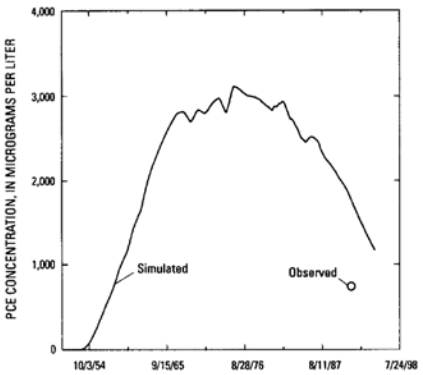

Figure F13. Simulated and observed tetrachloroethylene (PCE) concentrations at local water-supply well RW2, near ABC One-Hour Cleaners, Jacksonville, North Carolina, January 1951–December 1994 (see Figure F6 for location).

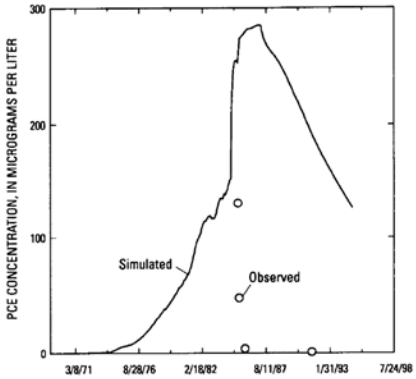

Figure F14. Simulated and observed tetrachloroethylene (PCE) concentrations at water-supply well TT-23, Tarawa Terrace, U.S. Marine Corps Base Camp Lejeune, North Carolina, January 1969–December 1994 (see Figure F6 for location).

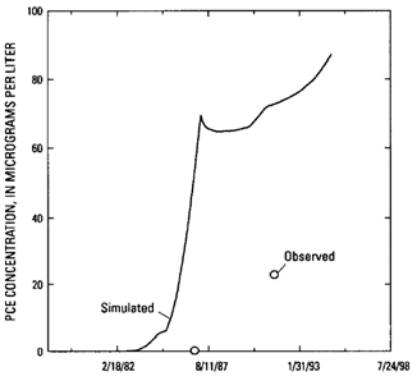

Figure F15. Simulated and observed tetrachloroethylene (PCE) concentrations at water-supply well TT-25, Tarawa Terrace, U.S. Marine Corps Base Camp Lejeune, North Carolina, January 1978–December 1994 (see Figure F6 for location).

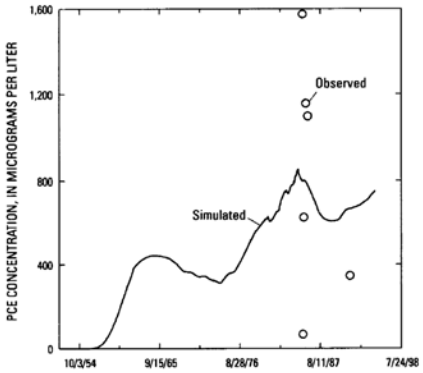

Figure F16. Simulated and observed tetrachloroethylene (PCE) concentrations at water-supply well TT-26, Tarawa Terrace, U.S. Marine Corps Base Camp Lejeune, North Carolina, January 1952–December 1994 (see Figure F6 for location).

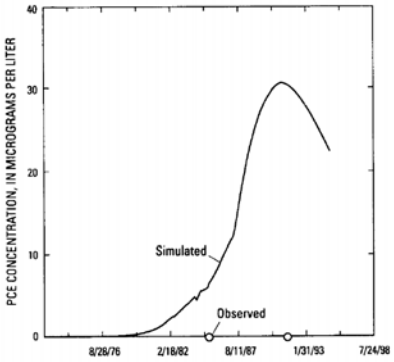

Figure F17. Simulated and observed tetrachloroethylene (PCE) concentrations at water-supply well TT-54, Tarawa Terrace, U.S. Marine Corps Base Camp Lejeune, North Carolina, January 1970–December 1994 (see Figure F6 for location).