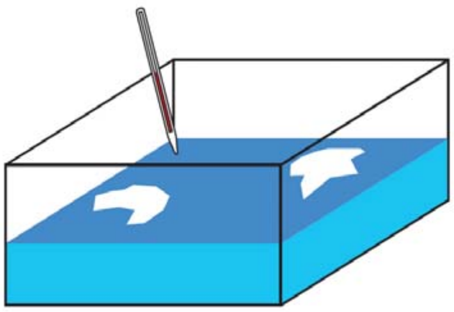
Data assimilation

Data assimilation provides a best estimate of the state of a system given a model of the system (and its error) and given observations (and their errors). The result is dynamically consistent and spatially complete (because of the use of a model) and close to observations whereever and whenever they are available, physicall realistic elsewhere.

How does it work? Let us consider the simplemost case: In front of us is a tank of water with melting ice. How warm is the water? According to theory, let us call this x_b, melting water should be near 0 °C. We also have a thermometer and measure the temperature, we call this: y. The thermometer indicates 0.2 °C. So we have two sources of information: a simple model and an observation. Both have their errors. Our model only holds for pure water, but our water might not be pure, for instance. Let us assume, the standard deviation of the error (σ_b) is 0.5 °C. Likewise, also the thermometer is perhaps not accurate. Let us assume an standard error (σ_y) of 0.2 °C for the thermometer.



"Model" $x_b = 0.0 \pm 0.5°C$

(theory; ice is melting)

Observation $y = 0.2 \pm 0.2°C$

We can now formulate a cost function. The best estimate x (called „analysis“) of the temperature of the water in the tank (for Gaussian error it can be shown that this is optimal) is the one that minimizes the following function J:

$$J(x) = \frac{(x - x_b)^2}{\sigma_b^2} + \frac{(y - x)^2}{\sigma_y^2}$$

This cost function has an analytic solution

$$x = \frac{\dfrac{x_b}{\sigma_b^2} + \dfrac{y}{\sigma_y^2}}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_y^2}} = x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_y^2}(y - x_b)$$

Note that the first expression is nothing else than a weighted error: models and observations are weighted by the inverse of their respective standard errors. The same weighted average can also be formulated as an update to the model (formulation on the right hand side), where the update is proportional to the misfit between model and observation (and proportional to the ratio of the model error variance to the sum of the error variances). In the concrete example, the solution is 0.1724 °C.