APPENDIX A *Estimation of "Data Error"*

There is always some noise or "data error" due to the amount of contributed data in the misfit objective function computation. The data error contribution could be estimated assuming there is (ε^m) due to insufficient parameterisation of models and the variance of the model error (σ^2_m) is zero. We also assume that there is uncorrelated random Gaussian noise (ε^d) in the data with zero mean and the standard deviation of ($\sigma^2_d=\sigma^2$). We start the meaning of the misfit, ***M***, as below:

$$M = \frac{\sum_{i=1}^{N} (y_i^{obs} - y_i^{mod})^2}{\sigma^2} \tag{B.1}$$

where ***N*** is number of data points, ***y*** with super scripts ***obs*** and ***mod*** refer to observed and model responses, respectively. We then define:

$$y_i^{obs} = y_i^{truth} + \varepsilon_i^d \tag{B.2}$$

$$y_i^{obs} = y_i^{truth} + \varepsilon_i^m \tag{B.3}$$

and

$$\sigma^2 = \frac{\sum (\varepsilon_i^d)^2}{N} \tag{B.4}$$

where ***truth*** refers to true response. When we assume the model is as truth (i.e. $\varepsilon^m = 0$), we could substitute the above relations in the Equation (B.1). We would obtain:

$$\boldsymbol{M = N} \qquad \text{(B.6)}$$

which represents the "data error" in the misfit, if we ignore the model error. This is also called "misfit of the noise". Nevertheless we typically keep the model error in Equation B.1, and assume that the noise in the data and model error are not correlated. Then for:

$$\frac{2\sum \varepsilon_i^m . \varepsilon_i^d}{\sigma^2} \approx 0.0 \qquad \text{(B.7)}$$

We obtain :

$$M \approx N + \frac{\sum \varepsilon_m^2}{\sigma} \qquad \text{(B.8)}$$

In Equation (B.8), $\frac{\sum \varepsilon_m^2}{\sigma}$ represents the misfit of the "parameter error".