This idea arose from a conversation that I had with Zhongmin Tao and Aibing Li at the AGU.

Consider an inversion that involves $M$ model parameters $\mathbf{m}_b$ that are constrained by two sets of data equations:

\[ \mathbf{g}_v(\mathbf{m}_b) = \mathbf{d}_v \quad \text{with} \quad \sigma^2_{d_v} \]
\[ \mathbf{g}_h(\mathbf{m}_b) = \mathbf{d}_h \quad \text{with} \quad \sigma^2_{d_h} \]

A common problem with such an inversion is that no reasonable $\mathbf{m}_b$ may satisfy both data equations, because: (a) the theories $\mathbf{g}_v(\mathbf{m})$ and $\mathbf{g}_h(\mathbf{m})$ may not be quite right; or (b) the noise in the data $\mathbf{d}_v$ and $\mathbf{d}_h$ might be mischaracterized. In such a case it may be advantageous to consider each data equation to depend on a different set of model parameters, labeled $v$ and $h$:

\[ \mathbf{g}_v(\mathbf{m}_v) = \mathbf{d}_v \quad \text{with} \quad \sigma^2_{d_v} \]
\[ \mathbf{g}_h(\mathbf{m}_h) = \mathbf{d}_h \quad \text{with} \quad \sigma^2_{d_h} \]

The overall vector model parameters $\mathbf{m} = [\mathbf{m}_v, \mathbf{m}_h]^T$ is of length $2M$. The notion that $\mathbf{m}_v = \mathbf{m}_h$ can then be implemented as prior information with variance $\sigma^2_{\ell}$. The quality of the assumption $\mathbf{m}_v = \mathbf{m}_h$ can then be evaluated with a standard squeezing test.

The linearized generalized least squares formulation is derived as follows:

(A) The $N_v + N_h$ data equations are linearized about trial solutions $\mathbf{m}_v^0$ and $\mathbf{m}_h^0$ with unknown corrections $\Delta \mathbf{m}_v$ and $\Delta \mathbf{m}_h$ via matrices $[\mathbf{G}_v]_{ij} = \partial g_{vi}/\partial m_{vj}|_{m_v^0}$ and $[\mathbf{G}_h]_{ij} = \partial g_{hi}/\partial m_{hj}|_{m_h^0}$ of partial derivatives:

\[ \mathbf{g}_v(\mathbf{m}_v^0) + \mathbf{G}_v \Delta \mathbf{m}_v = \mathbf{d}_v \quad \text{with} \quad \sigma^2_{d_v} \]
\[ \mathbf{g}_h(\mathbf{m}_h^0) + \mathbf{G}_h \Delta \mathbf{m}_h = \mathbf{d}_h \quad \text{with} \quad \sigma^2_{d_h} \]

(B) The $2M$ regularization equations for closeness to a base (prior) model $\mathbf{m}_b$ are:

\[ \mathbf{m}_v^0 + \Delta \mathbf{m}_v = \mathbf{m}_b^v \quad \text{with} \quad \sigma^2_{b_v} \]
\[ \mathbf{m}_h^0 + \Delta \mathbf{m}_h = \mathbf{m}_b^h \quad \text{with} \quad \sigma^2_{b_h} \]

(C) The $2K$ regularization equations for the smoothness of the solution, with first (or second) derivative operator $\mathbf{D}$, are:
The regularization equations for the $v$ and $h$ models being close to one another are:

\[
D_v(m_v^0 + \Delta m_v) = 0 \quad \text{with} \quad \sigma_{v}^2
\]
\[
D_h(m_h^0 + \Delta m_h) = 0 \quad \text{with} \quad \sigma_{h}^2
\]

Thus, the generalized least squares equations are:

\[
F \begin{bmatrix} \Delta m_v \\ \Delta m_h \end{bmatrix} = \begin{bmatrix}
\sigma_{d_v}^{-2} G_v & 0 \\
0 & \sigma_{d_h}^{-2} G_h \\
\sigma_{b_v}^{-2} I & 0 \\
0 & \sigma_{b_h}^{-2} I \\
\sigma_{a_v}^{-2} D_v & 0 \\
0 & \sigma_{a_h}^{-2} D_h \\
\sigma_c^{-2} I & -\sigma_c^{-2} I \\
\end{bmatrix} \begin{bmatrix} \Delta m_v \\ \Delta m_h \end{bmatrix} = \begin{bmatrix}
\sigma_{d_v}^{-2} (d_v - g_v(m_v^0)) \\
\sigma_{d_h}^{-2} (d_h - g_h(m_h^0)) \\
\sigma_{b_v}^{-2} (m_v^b - m_v^0) \\
\sigma_{b_h}^{-2} (m_h^b - m_h^0) \\
-D_v m_v^0 \\
-D_h m_h^0 \\
\sigma_c^{-2} (-m_v^0 + m_h^0) \\
\end{bmatrix} = f
\]

and the generalized least squares solution is:

\[
\begin{bmatrix} \Delta m_v \\ \Delta m_h \end{bmatrix} = [F^T F]^{-1} F^T f
\]

This solution must be iterated, $m_v^0 \rightarrow m_v^0 + \Delta m_v$ with $g_v(m_v^0)$, $g_h(m_h^0)$, $G_v$ and $G_h$ being recomputed at the start of each iteration.

The variance $\sigma_c^2$ controls the degree to which the $v$ and $h$ solutions are forced to equal one another. They become more and more similar as $\sigma_c^2 \rightarrow 0$. The proposition that two data equations are incompatible, and require the different model parameters, can be tested by generating a series of solutions, each for a different $\sigma_c$ and examining the behavior of the prediction error

\[
E = \sigma_{d_v}^{-2} ||e_v||_2^2 + \sigma_{d_v}^{-2} ||e_v||_2^2 \quad \text{with} \quad e_v = d_v - g_v(m_v) \quad \text{and} \quad e_h = d_h - g_h(m_h)
\]

now viewed as $E(\sigma_c)$. Is the error significantly smaller for large $\sigma_c$ than for small $\sigma_c$?

The overall problem has $2M$ unknowns and $N = (N_v + N_h + 3M + 2K)$ constraints, so the total number of degrees of freedom are $N = N - 2M$. The data equations have approximately $\nu_E = N_v(N_v + N_h)/N$ degrees of freedom. The prediction error $E$ is chi-squared distributed with approximately $\nu_E$ degrees of freedom, and has mean $\nu_E$ and variance $2\nu_E$. An F-test can be used to test against the Null Hypothesis that the difference between $E(\sigma_c^{small})$ and $E(\sigma_c^{large})$ is due to random variation (as contrasted to the data requiring two different models). Only when the Null Hypothesis can be rejected to greater than 95% confidence can the data be said to require two different models.