Consider a linear inverse problem $Gm = d$ for data $d$ (of length $N$) and model parameters $m$ (of length $M$). The estimated model parameters are $m^{\text{est}} = G^{-g}d^{\text{obs}}$, where $G^{-g}$ is a generalized inverse, and their covariance is $C$. The problem that I am considering is how to select the subset $\hat{d}$, of length $\hat{N} < N$, that leads to model parameters $\hat{m}^{\text{est}}$ having a covariance $\hat{C}$ that is as close as possible to $C$.

The predicted data is $d^{\text{obs}} = Gm^{\text{est}}$. Inserting $m^{\text{est}} = G^{-g}d^{\text{obs}}$ into this equation yields $d^{\text{pre}} = GG^{-g}d^{\text{obs}} \equiv Nd^{\text{obs}}$. When data resolution matrix $N \equiv GG^{-g}$ is unequal to the identity matrix, each predicted data is a non-trivial linear combination of the observed data. The importance vector $n \equiv \text{diag}(N)$ quantifies the extent to which a datum contributes to its own prediction.

My idea is to solve the problem with all $N$ data, compute the importance $n$, and then remove the datum with the least importance, leading to a dataset of length $N - 1$. The process is repeated until the length $\hat{N}$ is reached.

The method seems to work on test cases of a strait line fit and a quadratic fit. Some theoretical development is going to be needed to understand the general case.
Example 1. Straight line $d_i = m_1 + x_i m_2$, where $x$ is an auxiliary variable, $N = 101$, each datum has a different prior variance $\sigma_{d_i}^2$, and $\hat{N} = 6$. Note in Figure 1 that the procedure selects low-error data from the ends of the interval.

Fig. 1. (A) The observed data $d_{\text{obs}}$ (black circles), the predicted data $d_{\text{pre}}$ (black line) and the predicted data (green line) based on the winnowed data $d_{\text{obs}}$ (red circles). (B) Square root of the prior variance of $d_{\text{obs}}$. (C) and (D) Square root of the posterior variances of $m_1$ and $m_2$, respectively, as a function of $\hat{N}$. 
Example 1. Same as the previous example, but for the quadratic curve $d_i = m_1 + x_i m_2 + x_i^2 m_3$. The solution is by weighted least-squares. Note in Figure 1 that the procedure selects low-error data from the ends of the interval and at its center.

Fig. 1. (A) The observed data $d_{\text{obs}}$ (black circles), the predicted data $d_{\text{pre}}$ (black line) and the predicted data (green line) based on the winnowed data $d_{\text{obs}}$ (red circles). (B) Square root of the prior variance of $d_{\text{obs}}$. (C), (D) and (E) Square root of the posterior variances of $m_1$, $m_2$ and $m_3$, respectively, as a function of $\tilde{N}$. 
% weighted least squares inversion
W = diag(sigmad.^(-2));
GMG = (G'*W*G)\(G'*W);
mest = GMG*dobs;
dpre = G*mest;
Cm = GMG * diag(sigmad.^2) * GMG';
NN = G * GMG; % data resolution matrix
n = diag(NN); % data importance

% set up for "whittling away" iteration
Nshort = N;
xshort = x;
dshort = dobs;
sigmadshort = sigmad;
mshort = mest;
NNshort = NN;
nshort = n;
Cmshort = Cm;
Cmall = zeros(N,M,M);
Cmall(1,:,:)=Cm;
fprintf('%d m %.2f %.2f cov slope %.4f
', Nshort, mshort(1), mshort(2), Cmshort(2,2));

% "whittling away" iteration
for iter = [2:N-Nfinal+1]
    [nsort, irow] = sort(nshort, 'descend');
    Nshort = Nshort-1;
irow = irow(1:Nshort);
dshort = dshort(irow);
xshort = xshort(irow);
sigmadshort = sigmadshort(irow);
Gshort = [ones(Nshort,1), xshort];
Wshort = diag(sigmadshort.^(-2));
GMGshort = (Gshort'*Wshort*Gshort)\(Gshort'*Wshort);
mshort = GMGshort*dshort;
dpreshort = Gshort*mshort;
Cmshort = GMGshort * diag(sigmadshort.^2) * GMGshort';
Cmall(iter,:,:)=Cmshort;
NNshort = Gshort * GMGshort; % data resolution matrix
nshort = diag(NNshort); % data importance
fprintf('%d m %.2f %.2f cov slope %.4f
', Nshort, mshort(1), mshort(2), Cmshort(2,2));
end