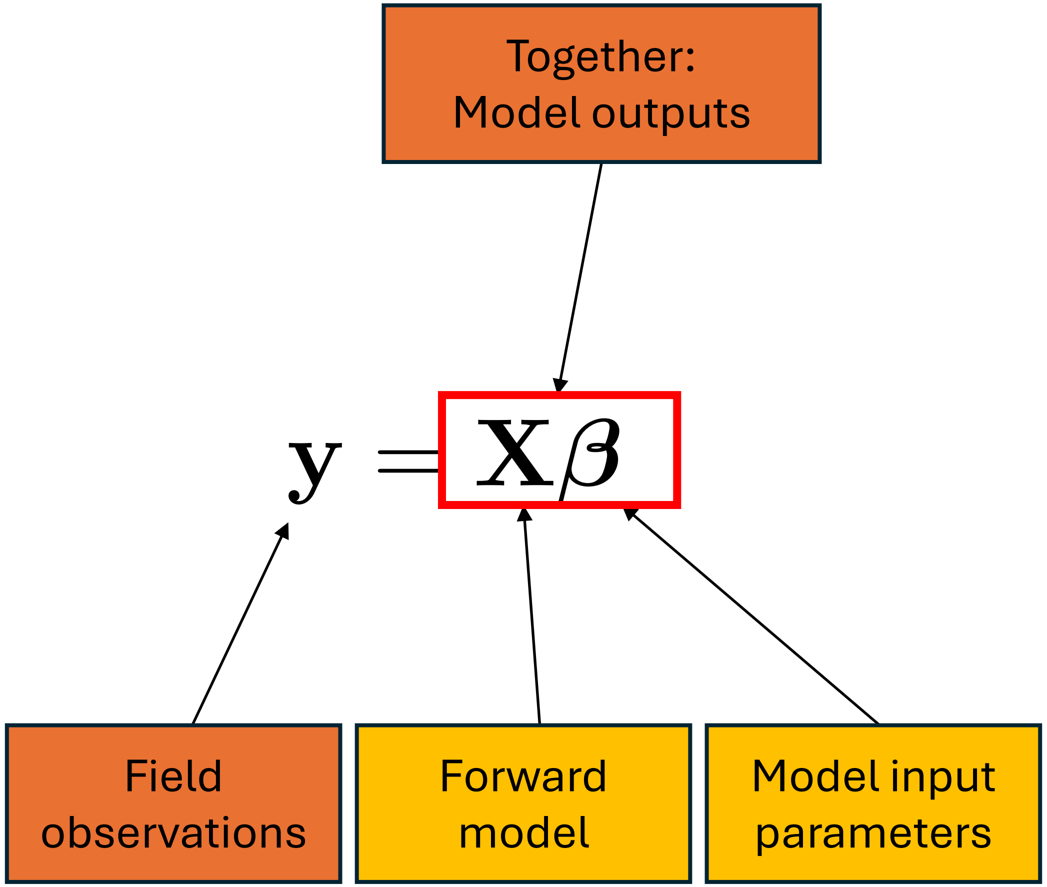
A little maths journey behind the GLM algorithm



$$X_{ij} = \frac{\partial y_i}{\partial \beta_j} \approx \frac{y_i\left(\beta_j + \Delta\beta_j\right) - y_i\left(\beta_j\right)}{\Delta\beta_j}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

Errors

Field observations

Forward model

Model input parameters

rearranging....

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

So we should minimize these errors right?

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

For nice mathy reasons, easier to minimize the square,
so let's define an **objective function** as the "sum of squared errors"

$$\Phi\left(\boldsymbol{\beta}\right) = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) = \sum_{i=1}^{N_{obs}}\left(y_i - \sum_{j=1}^{N_{par}} X_{ij}\beta_j\right)^2$$

But....should each observation be reproduced with equal quality by the model?

$$\Phi\left(\boldsymbol{\beta}\right) = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T\mathbf{Q}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) = \sum_{i=1}^{N_{obs}}\left(\frac{y_i - \sum_{j=1}^{N_{par}} X_{ij}\beta_j}{\sigma_i}\right)^2$$

How about a simple groundwater-based example?

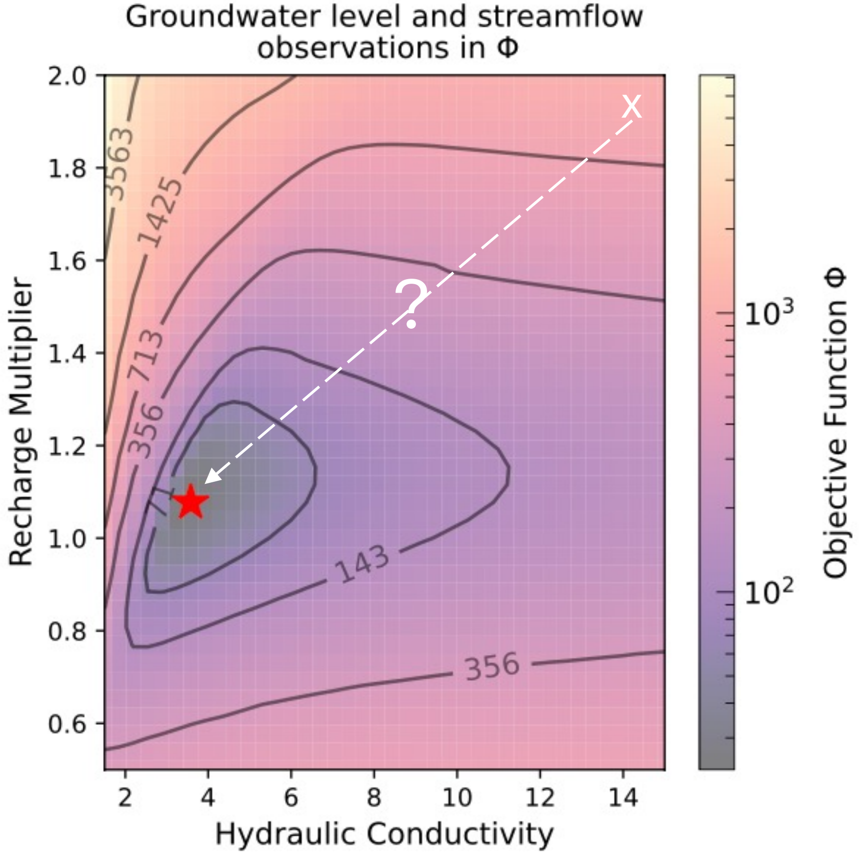
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

Groundwater levels and Streamflows ....

MODFLOW GW model

K and R

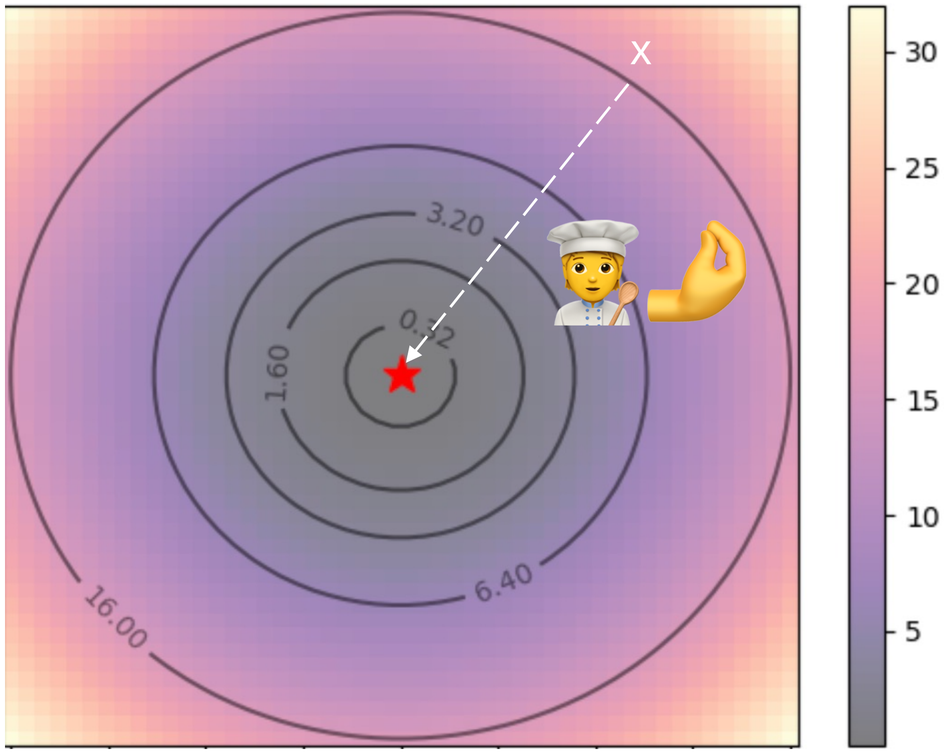
$$\Phi\left(\boldsymbol{\beta}\right) = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^{T}\mathbf{Q}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)$$



How can we find the minimum of this?

$$\Phi\left(\boldsymbol{\beta}\right) = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T \mathbf{Q}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)$$

Take the derivative and set it to zero



$$\frac{\partial \Phi\left(\hat{\boldsymbol{\beta}}\right)}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

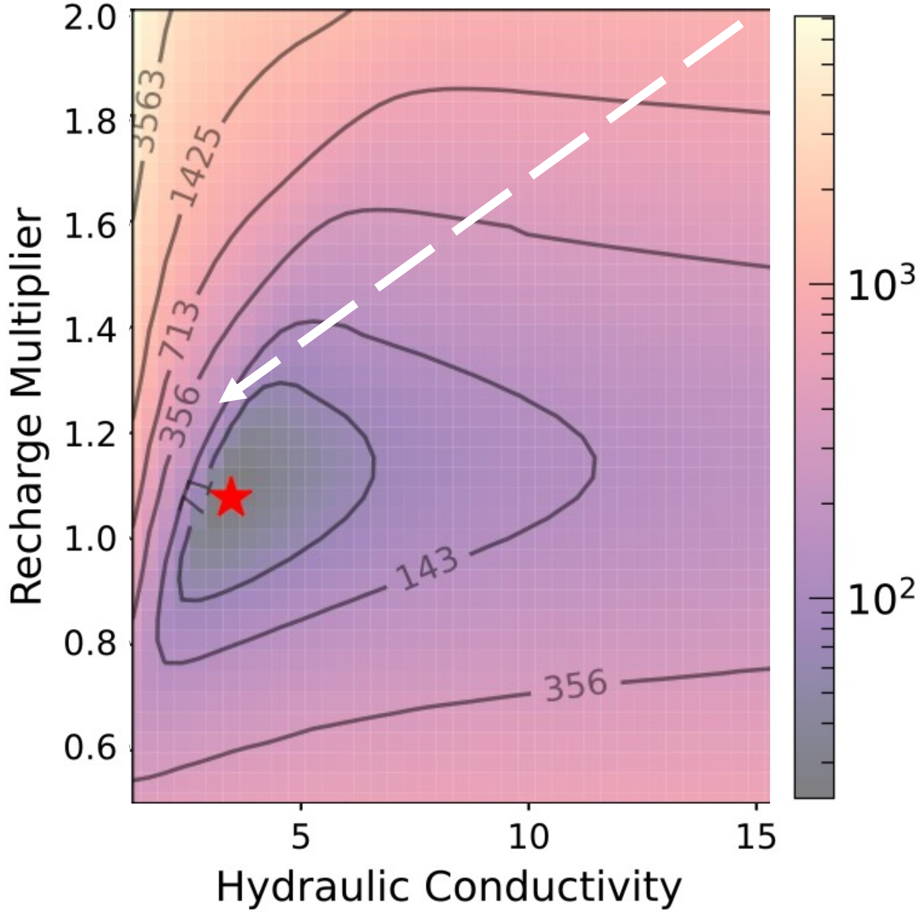
...through the magic of algebra (thanks Al-Khwarizmi!)

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}}$$

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}\mathbf{y}}$$

Al-Khwarizmi

But this is usually nonlinear

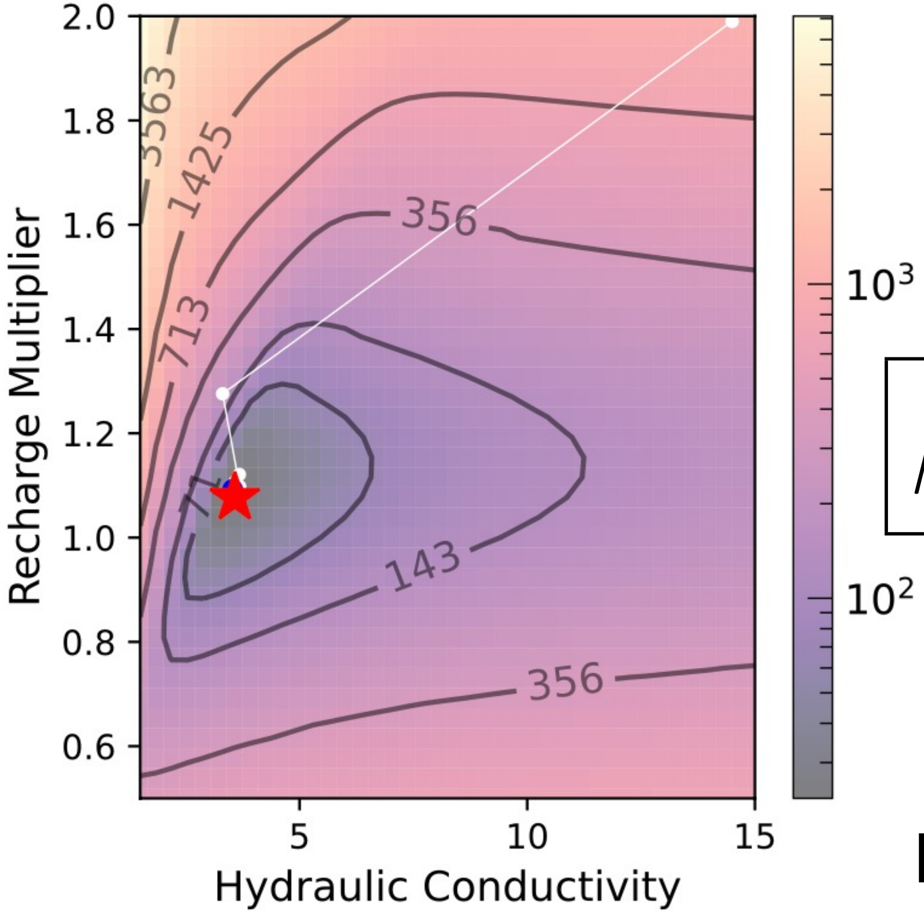


$$\hat{\beta} = \boxed{(\mathbf{X}^T\mathbf{Q}\mathbf{X})}^{-1}\boxed{\mathbf{X}^T}\mathbf{Q}\mathbf{y}$$

Curvature
(2^{nd} derivative)

Gradient/Slope
(1^{st} derivative)

So......let's iterate!



$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_{i+1} = \boxed{\hat{\boldsymbol{\beta}}_i} + \boxed{\left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}}\boxed{\left(\mathbf{y} - f\left(\boldsymbol{\beta}_i\right)\right)}$$

Previous parameters

Upgrade

Current misfit

Can we improve this?

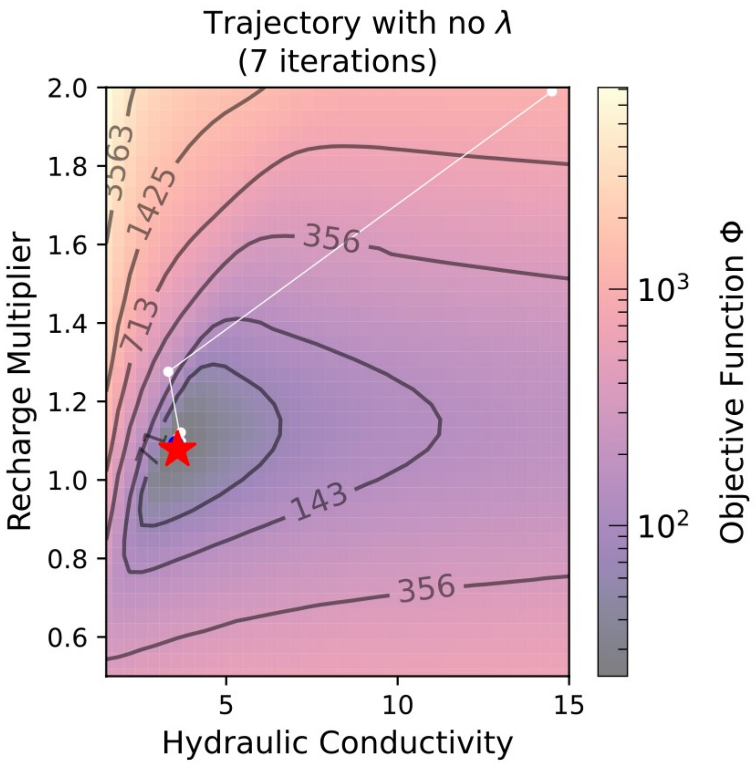
$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + \boxed{\left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)}^{-1} \boxed{\mathbf{X}^T}\mathbf{Q}\left(\mathbf{y} - f\left(\boldsymbol{\beta}_i\right)\right)$$

Curvature
(2nd derivative)

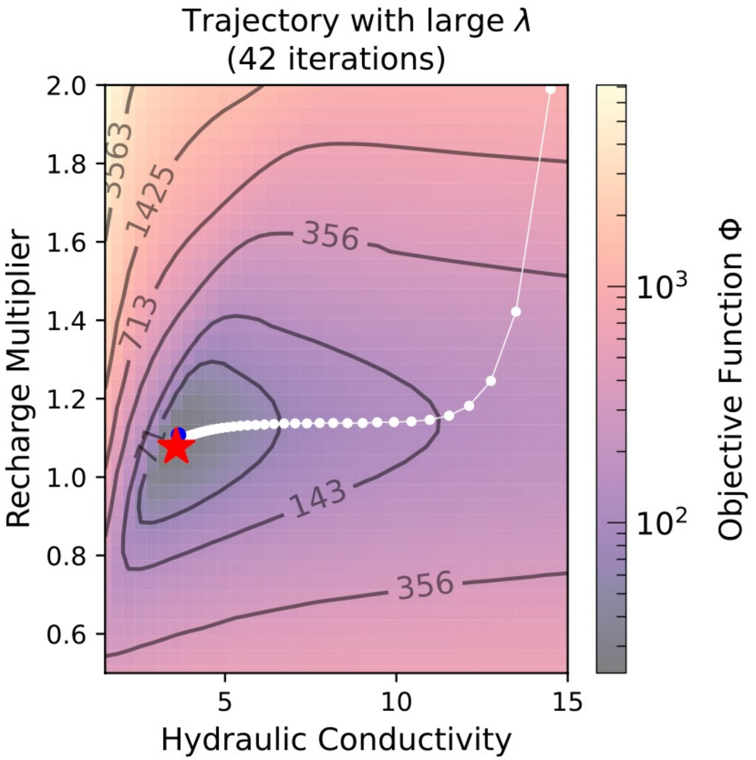
Gradient/Slope
(1st derivative)

$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + \left(\mathbf{X}^T\mathbf{Q}\mathbf{X} + \boxed{\lambda diag\left(\mathbf{X}^T\mathbf{X}\right)}\right)^{-1} \mathbf{X}^T\mathbf{Q}\left(\mathbf{y} - f\left(\boldsymbol{\beta}_0\right)\right)$$

Lambda – a way to adjust
the tradeoff of slope
And curvature

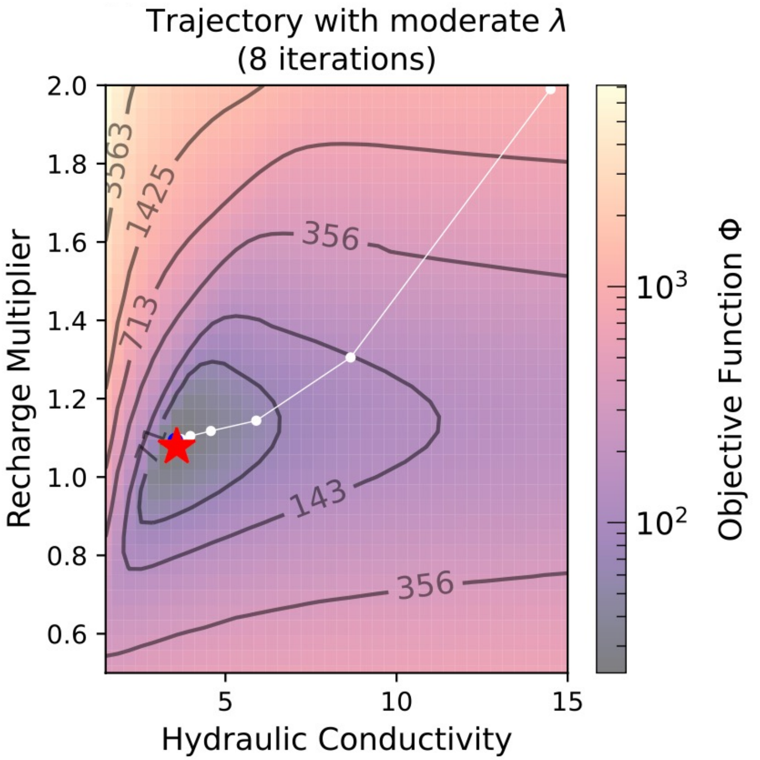


$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + \left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}\left(\mathbf{y} - f\left(\boldsymbol{\beta}_i\right)\right)$$



Knocks out the Curvature.
All gradient.
Small steps.

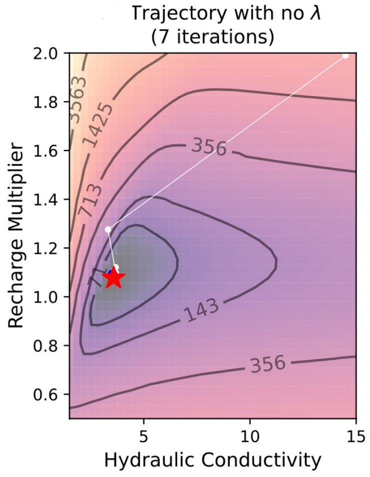
$$\left(\mathbf{X}^T\mathbf{Q}\mathbf{X} + \lambda diag\left(\mathbf{X}^T\mathbf{X}\right)\right)^{-1} \rightarrow$$
$$\left(\cancel{\mathbf{X}^T\mathbf{Q}\mathbf{X}} + \lambda diag\left(\mathbf{X}^T\mathbf{X}\right)\right)^{-1}$$

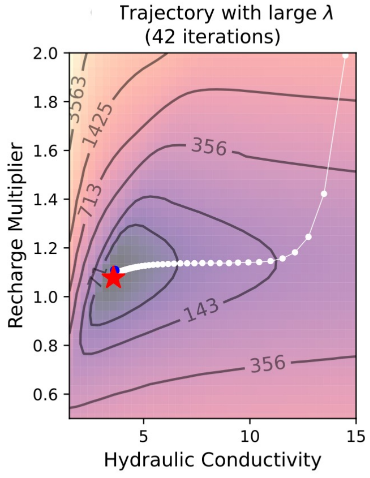


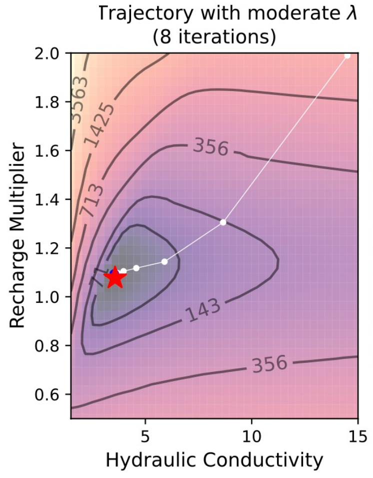
Nice balance.
Happy medium.
World peace.

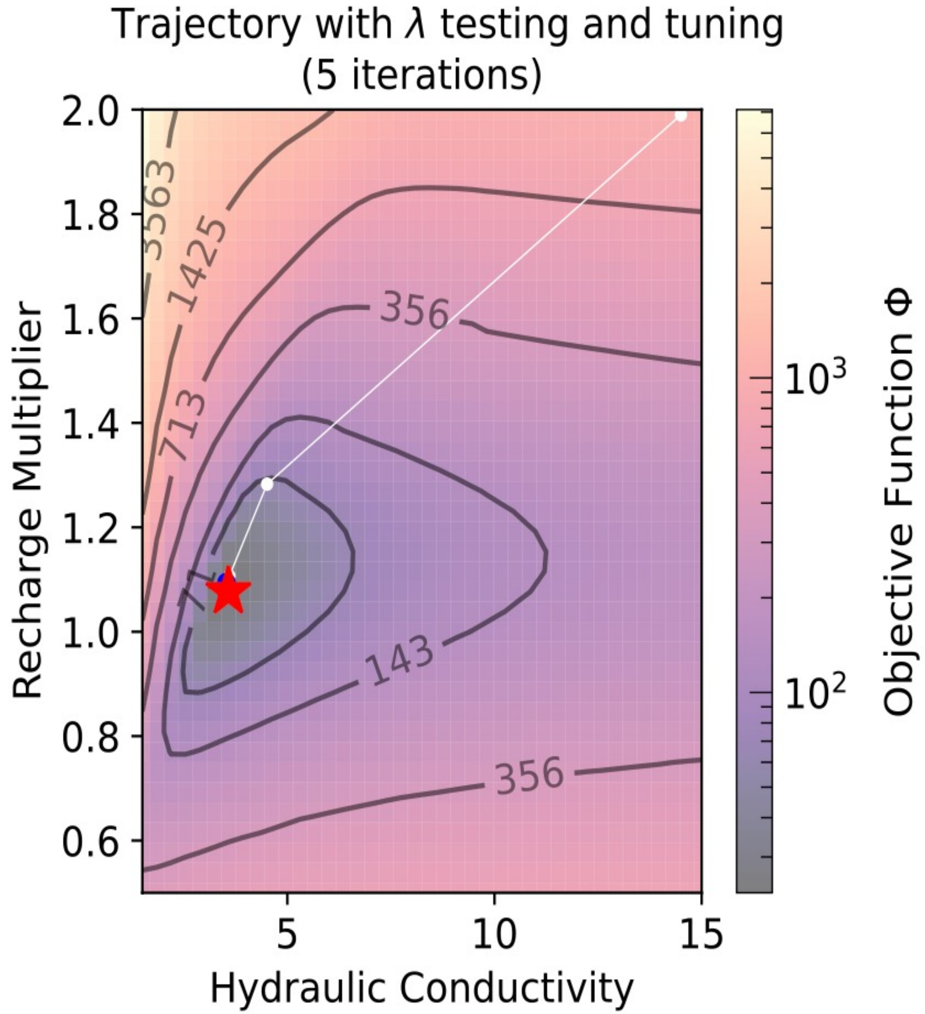
$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + \left(\cancel{\mathbf{X}^T\mathbf{Q}\mathbf{X}} + \lambda diag\left(\mathbf{X}^T\mathbf{X}\right)\right)^{-1}\mathbf{X}^T\mathbf{Q}\left(\mathbf{y} - f\left(\boldsymbol{\beta}_0\right)\right)$$

Sweet! So lambda is useful. What’s the right value?
We need heuristics – basically smart grid search









$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + \left(\mathbf{X}^T\mathbf{Q}\mathbf{X} + \lambda diag\left(\mathbf{X}^T\mathbf{X}\right)\right)^{-1}\mathbf{X}^T\mathbf{Q}\left(\mathbf{y} - f\left(\boldsymbol{\beta}_0\right)\right)$$

Want more details?

Paper in review at *Groundwater*

Fienen, White, Hayek (2024) *Methods Brief/ Parameter ESTimation with the Gauss Levenberg Marquardt algorithm: an intuitive guide*. Groundwater