An elicitation tool for prior beliefs about complex systems

Peter Craig and Michael Goldstein
(University of Durham)

References:

Craig, P.S., Goldstein, M., Seheult, A.H., Smith, J.A. (1997): Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments (with discussion), in Case studies in Bayesian statistics

Craig, P.S., Goldstein, M., Seheult, A.H., Smith, J.A. (1997) Constructing partial prior specifications for complex physical models (with discussion), in J Royal Statist Soc, series D.

[Thanks to EPSRC]

Problem: eliciting structured multivariate beliefs

We have a collection of quantities $y = (y_1, \ldots, y_p)$
(for example, quantities such as bottom hole pressures or oil production, at various wells)
whose values are at least partly determined by a collection $x = (x_1, \ldots, x_d)$.
(for example, quantities related to the geology of a reservoir, such as permeability of region or status of potential faults)

We want to elicit expert beliefs about the relationships between y and x
(so that we can change the collection of x quantities to optimise y, or , in our example, so that we can use observation on y to make deductions about x)

Characteristics of the problem

Typically, in the applications we are concerned with:
— the collections x and y are large
— there is a large amount of relevant prior knowledge for the elicitation
— the physical context for the various quantities is essential to the structuring of information

Implications

The size and complexity of the problem require that:
— the expert must be guided through a structured elicitation;
— this elicitation must be directed by appropriate computer software;
— this software should be based on a physical representation of the problem.

The software should:
— support qualitative structuring of prior beliefs to simplify and organise the elicitation;
— offer tools for individual quantifications of limited aspects of prior beliefs (in our case, first and second order prior structure);
— give feedback on the implications of the qualitative and quantitative elicitation.

The elicitation tool that we shall describe is based around certain forms of belief about the relationships between the quantities which we now describe.

An elicitation tool for prior beliefs about complex systems

Peter Craig and Michael Goldstein

(University of Durham)

References:

Craig, P.S., Goldstein, M., Seheult, A.H., Smith, J.A. (1997): Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments (with discussion), in Case studies in Bayesian statistics

Craig, P.S., Goldstein, M., Seheult, A.H., Smith, J.A. (1997) Constructing partial prior specifications for complex physical models (with discussion), in J Royal Statist Soc, series D.

[Thanks to EPSRC]

Problem: eliciting structured multivariate beliefs

We have a collection of quantities $y = (y_1, \ldots, y_p)$

(for example, quantities such as bottom hole pressures or oil production, at various wells)

whose values are at least partly determined by a collection $x = (x_1, \ldots, x_d)$.

(for example, quantities related to the geology of a reservoir, such as permeability of region or status of potential faults)

We want to elicit expert beliefs about the relationships between y and x

(so that we can change the collection of x quantities to optimise y, or , in our example, so that we can use observation on y to make deductions about x)

Characteristics of the problem

Typically, in the applications we are concerned with:

— the collections x and y are large
— there is a large amount of relevant prior knowledge for the elicitation
— the physical context for the various quantities is essential to the structuring of information

Implications

The size and complexity of the problem require that:

— the expert must be guided through a structured elicitation;
— this elicitation must be directed by appropriate computer software;
— this software should be based on a physical representation of the problem.

The software should:

— support qualitative structuring of prior beliefs to simplify and organise the elicitation;
— offer tools for individual quantifications of limited aspects of prior beliefs (in our case, first and second order prior structure);
— give feedback on the implications of the qualitative and quantitative elicitation.

The elicitation tool that we shall describe is based around certain forms of belief about the relationships between the quantities which we now describe.

Relating a single quantity y to a collection x

We shall first describe how we express our beliefs about the relationship between a single quantity y (such as the pressure at a particular oil well at a particular time) and the collection of quantities $x = (x_1, ..., x_d)$ (such as geological measures within a reservoir). We write y as $y(x)$. We choose a (small) subset $x^* = (x_0^*, ..., x_s^*)$ of the components of x (with $x_0^* = 1$) which we believe, a priori, to be important in explaining the variance in $y(x)$. We express this relation as

$$y(x) = \sum_{j,k=0}^{s} \beta_{jk} x_j^* x_k^* + \epsilon(x)$$

where β_{jk} are constants and $\epsilon(x^*)$ is a stationary stochastic process with mean 0 and variance structure

$$\frac{1}{2}\mathrm{Var}\big[\epsilon(u) - \epsilon(v)\big] = \sigma_0^2\big(1 - \delta(\tilde{u} - \tilde{v})\big) + \sum_{j=1}^{s} \sigma_j^2(1 - e^{-c_j|u_j^* - v_j^*|^2})$$

where $\tilde{x}$ denotes that components of x which are not in x^*.

(We here describe the simplest form of specification which is sufficient to demonstrate the function of the elicitation tool. In general we would:

- allow general functions of the variables x^* in the relation with $y(x)$;
- introduce additional terms into the structure ϵ to represent the residual variation in $y(x)$ corresponding to those elements in x but not x^*;
- introduce constraints on ϵ to force orthogonality with the linear terms in the relation.)

Eliciting beliefs for the structure

To specify second order beliefs for the above relationship, we must

- select the quantities x^*,
- specify prior means, variances and covariances for each β_{jk}
- specify the values $\sigma_0, \ldots, \sigma_s$ and $c_1, \ldots, c_s$.

We need to repeat this procedure for each individual quantity y.

Therefore, we first qualitatively structure beliefs, by identifying which variables y_i depend on the same collections x^*, and, more strongly, for which collections of y_i quantities our beliefs over β, ϵ are exchangeable. We then choose strategic quantities from our exchangeable groups, perform quantitative elicitations for these choices, and spread these assessments across all other members of the group.

Describing spatial correlation

The final stage in the elicitation is to describe the correlation structure between the y_i. In our case, this is strongly influenced by spatial considerations. We therefore describe how we may elaborate our description to incorporate spatial variation. Again, we describe the simplest case which illustrates the functioning of the tool.

Suppose that each y is measured at a spatial location $s = (s_1, s_2)$. Therefore, we write each quantity as $y(x, s)$ to indicate the dependence on both x and location s. Suppose, for simplicity, that over a particular region we judge that each $y(x, s)$ depends on the same choice x^*. We express this relationship as

$$y(x, s) = \sum_{j,k} \beta_{jk}(s) x_j^* x_k^* + \epsilon(x, s)$$

where $\beta_{jk}(s)$ are constants, related over s as

$$\beta_{jk}(s) = \beta_{jk0} + \beta_{jk1}s_1 + \beta_{jk2}s_2 + \alpha(s)$$

where

- β_{jki} are constants representing the differential variation in the two chosen directions
- $\alpha(s)$ is a mean zero stationary stochastic process with parameters to be specified,
- $\epsilon(x, s)$ is a mean zero stationary stochastic process which we decompose as

$$\epsilon(x, s) = \gamma(x^*) + \psi(x, s)$$

where $\gamma(x^*)$ is a stationary stochastic process in x^* which expresses the common variation in ϵ at each site; $\psi(x, s)$ represents individual site variation, so that $\mathrm{Cov}[\psi(u, s), \psi(v, s')] = 0$ when $s \neq s'$.

Often, we constrain our choice of parameters for the various stochastic processes for example, we may set $\mathrm{cov}(\gamma(u^*), \gamma(v^*)) = \phi_\epsilon \mathrm{cov}(\delta(u^*, s), \delta(v^*, s))$.

Elicitation for this structure follows by

- choosing a site s_0 at which to set $s_0 = 0$,
- eliciting beliefs for the univariate model at this site, namely mean and variance structure for each β_{jk0} and variance structure for $\epsilon(x^*, 0)$
- eliciting spatial variation, e.g. means and variances for β_{jk1}, β_{jk2} the parameters for α and the constant ϕ_ϵ.

Relating a single quantity y to a collection x

We shall first describe how we express our beliefs about the relationship between a single quantity y (such as the pressure at a particular oil well at a particular time) and the collection of quantities $x = (x_1, ..., x_d)$ (such as geological measures within a reservoir). We write y as $y(x)$. We choose a (small) subset $x^* = (x_0^*, ..., x_s^*)$ of the components of x (with $x_0^* = 1$) which we believe, a priori, to be important in explaining the variance in $y(x)$. We express this relation as

$$y(x) = \sum_{j,k=0}^{s} \beta_{jk} x_j^* x_k^* + \epsilon(x)$$

where β_{jk} are constants and $\epsilon(x^*)$ is a stationary stochastic process with mean 0 and variance structure

$$\frac{1}{2}\mathrm{Var}\big[\epsilon(u) - \epsilon(v)\big] = \sigma_0^2\big(1 - \delta(\tilde{u} - \tilde{v})\big) + \sum_{j=1}^{s} \sigma_j^2(1 - e^{-c_j|u_j^* - v_j^*|^2})$$

where $\tilde{x}$ denotes that components of x which are not in x^*.

(We here describe the simplest form of specification which is sufficient to demonstrate the function of the elicitation tool. In general we would:

- allow general functions of the variables x^* in the relation with $y(x)$;
- introduce additional terms into the structure ϵ to represent the residual variation in $y(x)$ corresponding to those elements in x but not x^*;
- introduce constraints on ϵ to force orthogonality with the linear terms in the relation.)

Eliciting beliefs for the structure

To specify second order beliefs for the above relationship, we must

- select the quantities x^*,
- specify prior means, variances and covariances for each β_{jk}
- specify the values $\sigma_0, \ldots, \sigma_s$ and $c_1, \ldots, c_s$.

We need to repeat this procedure for each individual quantity y.

Therefore, we first qualitatively structure beliefs, by identifying which variables y_i depend on the same collections x^*, and, more strongly, for which collections of y_i quantities our beliefs over β, ϵ are exchangeable. We then choose strategic quantities from our exchangeable groups, perform quantitative elicitations for these choices, and spread these assessments across all other members of the group.

Describing spatial correlation

The final stage in the elicitation is to describe the correlation structure between the y_i. In our case, this is strongly influenced by spatial considerations. We therefore describe how we may elaborate our description to incorporate spatial variation. Again, we describe the simplest case which illustrates the functioning of the tool.

Suppose that each y is measured at a spatial location $s = (s_1, s_2)$. Therefore, we write each quantity as $y(x, s)$ to indicate the dependence on both x and location s. Suppose, for simplicity, that over a particular region we judge that each $y(x, s)$ depends on the same choice x^*. We express this relationship as

$$y(x, s) = \sum_{j,k} \beta_{jk}(s) x_j^* x_k^* + \epsilon(x, s)$$

where $\beta_{jk}(s)$ are constants, related over s as

$$\beta_{jk}(s) = \beta_{jk0} + \beta_{jk1}s_1 + \beta_{jk2}s_2 + \alpha(s)$$

where

- β_{jki} are constants representing the differential variation in the two chosen directions
- $\alpha(s)$ is a mean zero stationary stochastic process with parameters to be specified,
- $\epsilon(x, s)$ is a mean zero stationary stochastic process which we decompose as

$$\epsilon(x, s) = \gamma(x^*) + \psi(x, s)$$

where $\gamma(x^*)$ is a stationary stochastic process in x^* which expresses the common variation in ϵ at each site; $\psi(x, s)$ represents individual site variation, so that $\mathrm{Cov}[\psi(u, s), \psi(v, s')] = 0$ when $s \neq s'$.

Often, we constrain our choice of parameters for the various stochastic processes for example, we may set $\mathrm{cov}(\gamma(u^*), \gamma(v^*)) = \phi_\epsilon \mathrm{cov}(\delta(u^*, s), \delta(v^*, s))$.

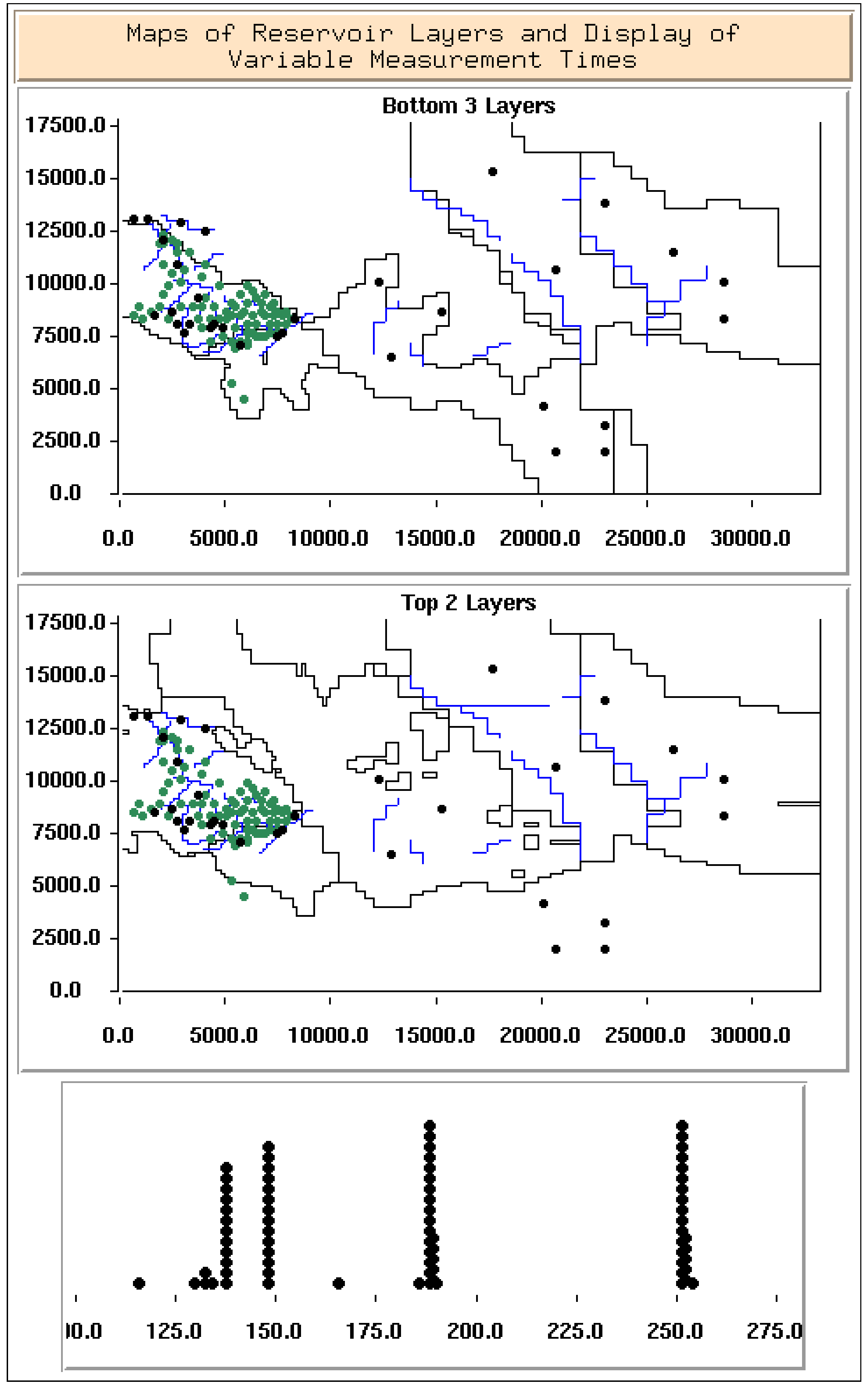
Elicitation for this structure follows by

- choosing a site s_0 at which to set $s_0 = 0$,
- eliciting beliefs for the univariate model at this site, namely mean and variance structure for each β_{jk0} and variance structure for $\epsilon(x^*, 0)$
- eliciting spatial variation, e.g. means and variances for β_{jk1}, β_{jk2} the parameters for α and the constant ϕ_ϵ.

The Elicitation Sequence

- Specify ranges and central values for x-variables.
- Organise y-variables into groups and subgroups.
- Specify active x-variables for each group and subgroup.
- Choose a “master” variable for each group
- Elicit beliefs about the effects of the chosen active x-variables for the “master” variable in each group.
- Where appropriate elicit spatial variation in the effects of active x-variables for a group.

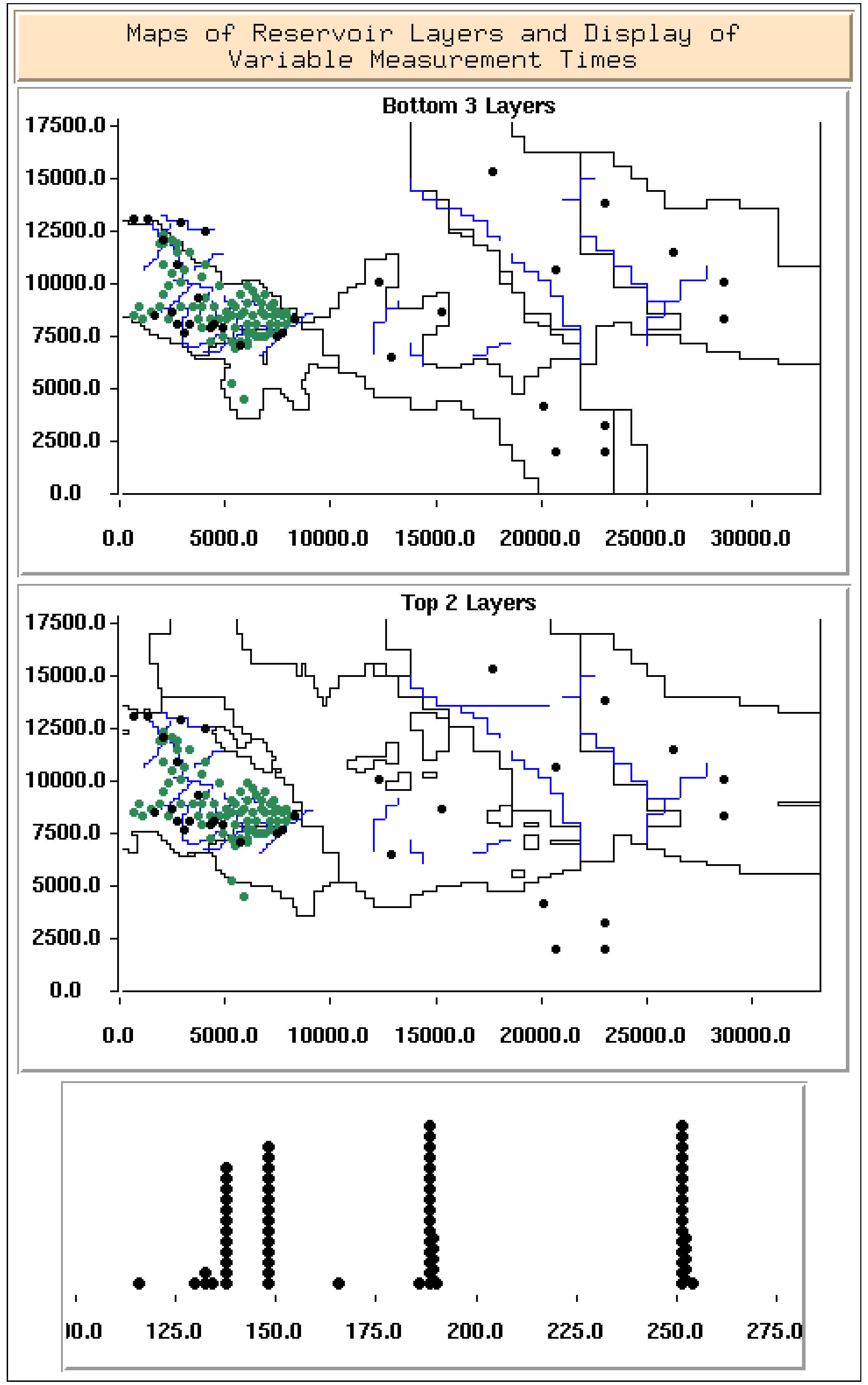
Metaphor for the System



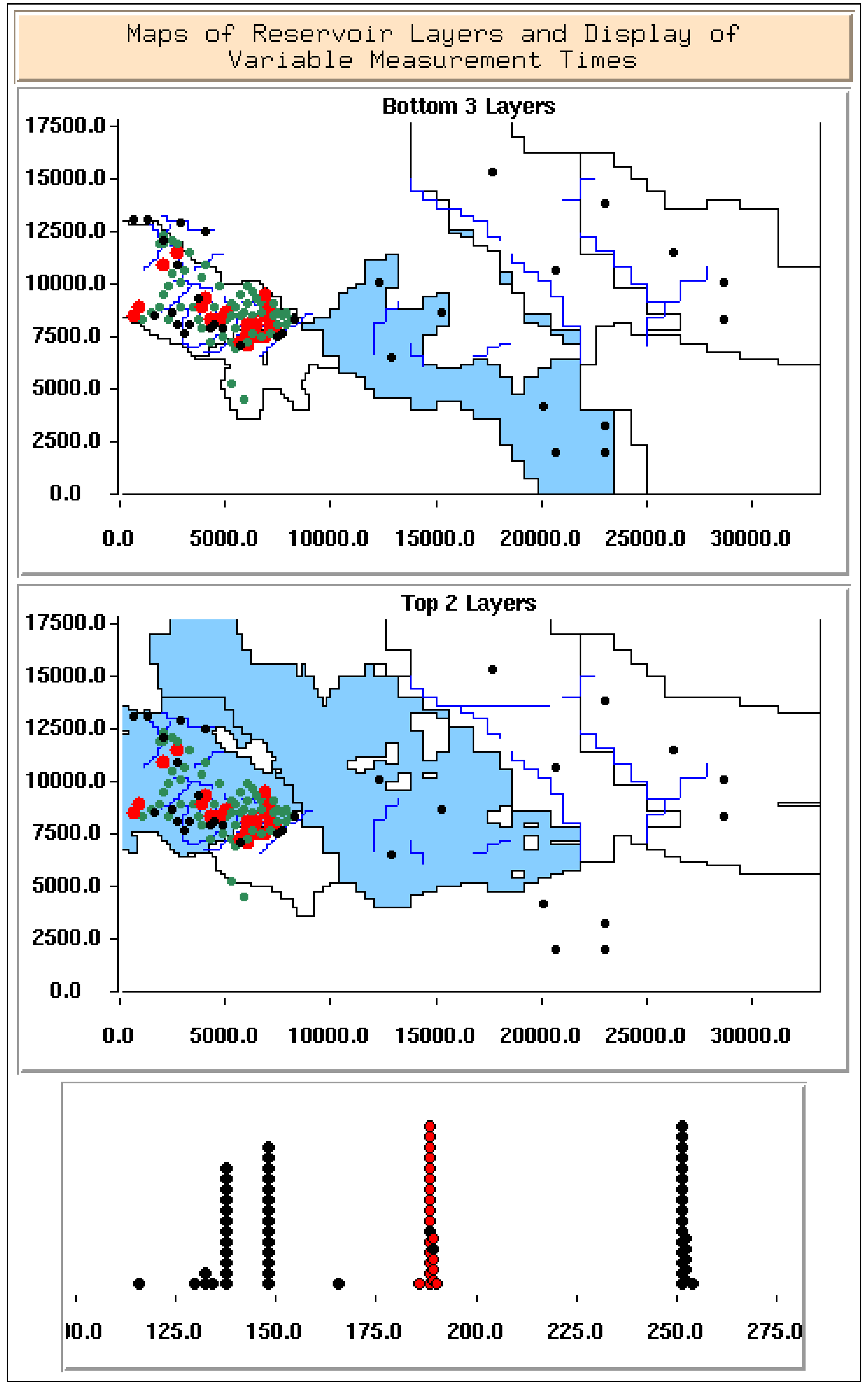
The Elicitation Sequence

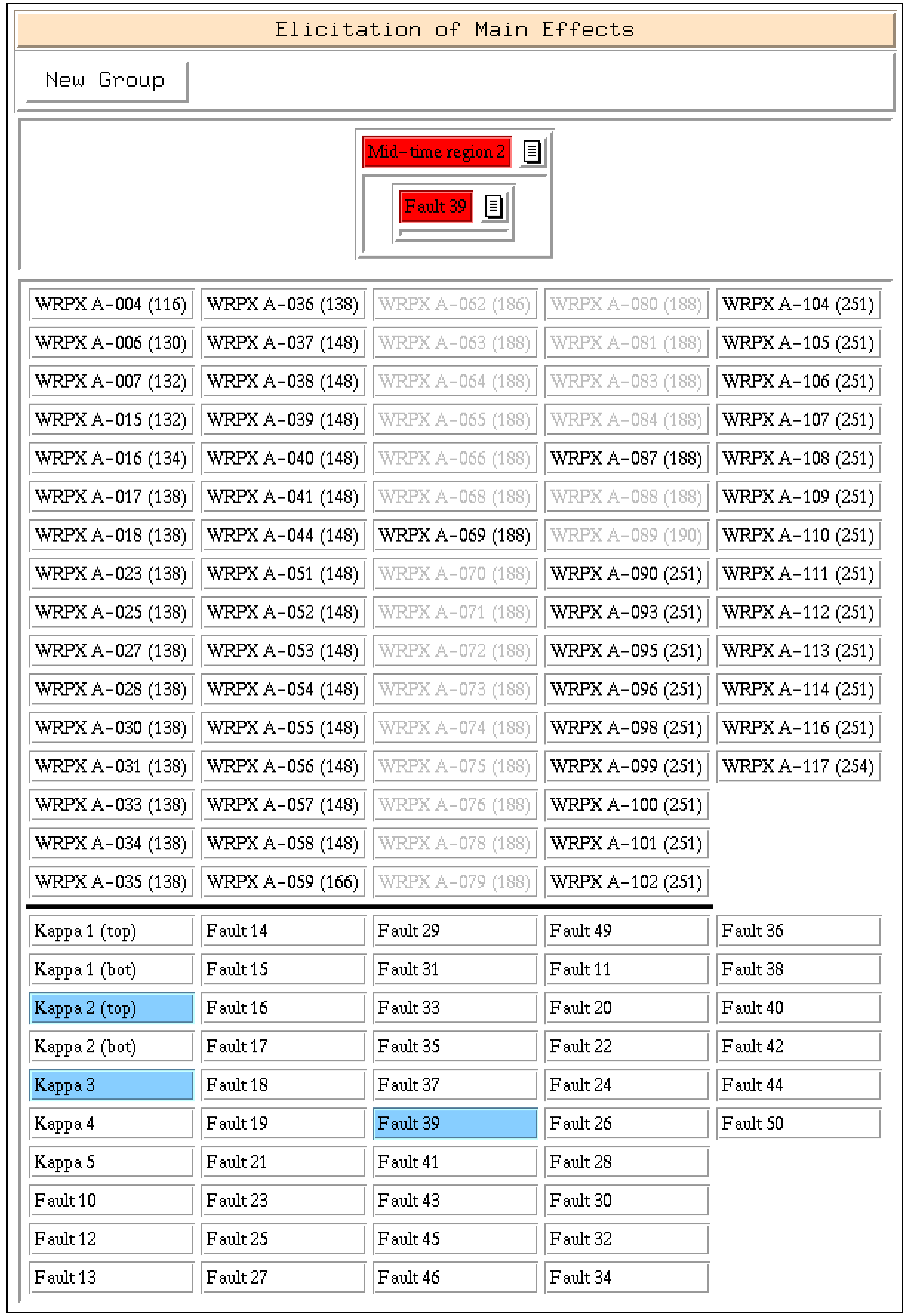
- Specify ranges and central values for x-variables.
- Organise y-variables into groups and subgroups.
- Specify active x-variables for each group and subgroup.
- Choose a "master" variable for each group
- Elicit beliefs about the effects of the chosen active x-variables for the "master" variable in each group.
- Where appropriate elicit spatial variation in the effects of active x-variables for a group.

Metaphor for the System

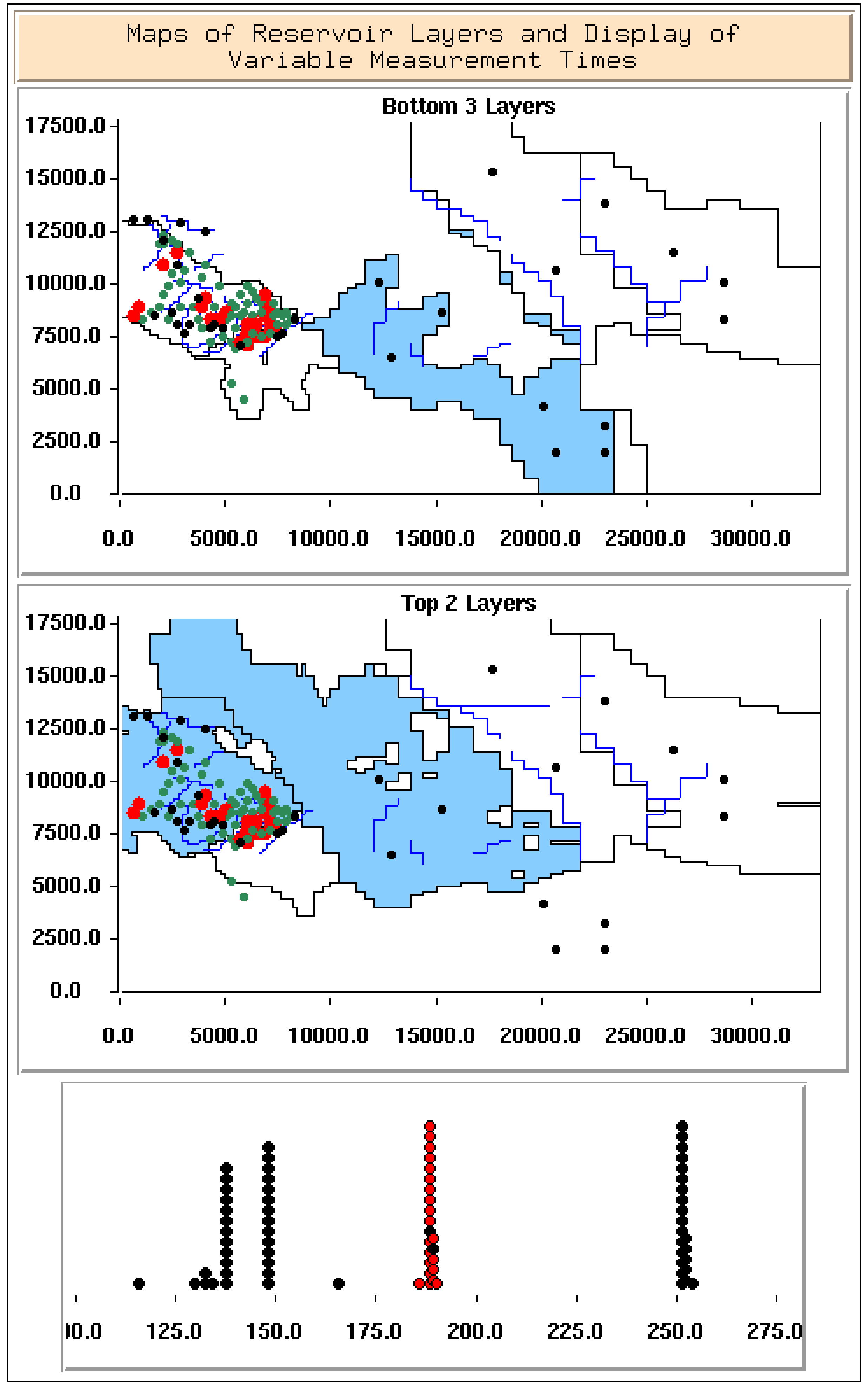


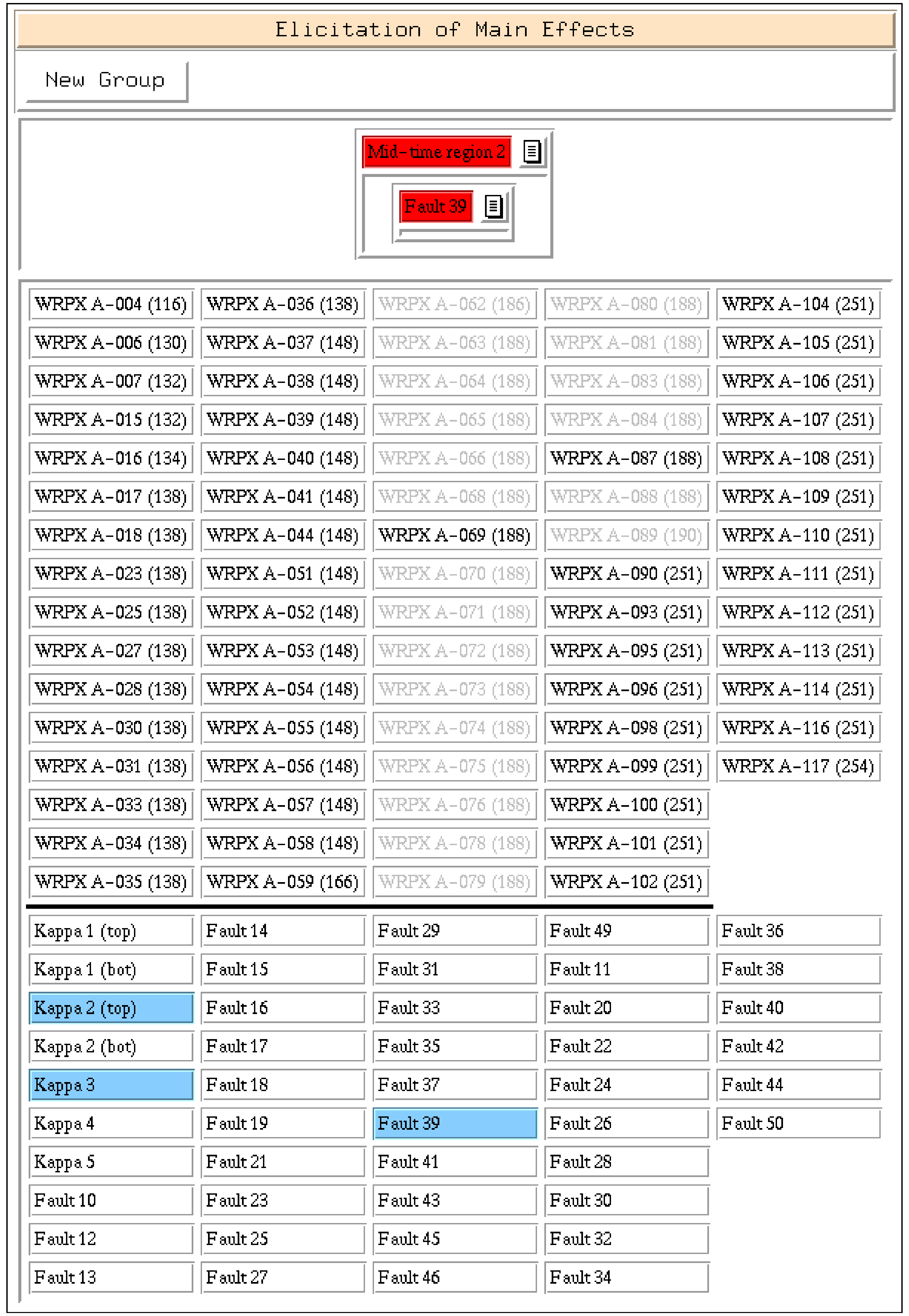
Construction of Groups and Choice of Active Variables



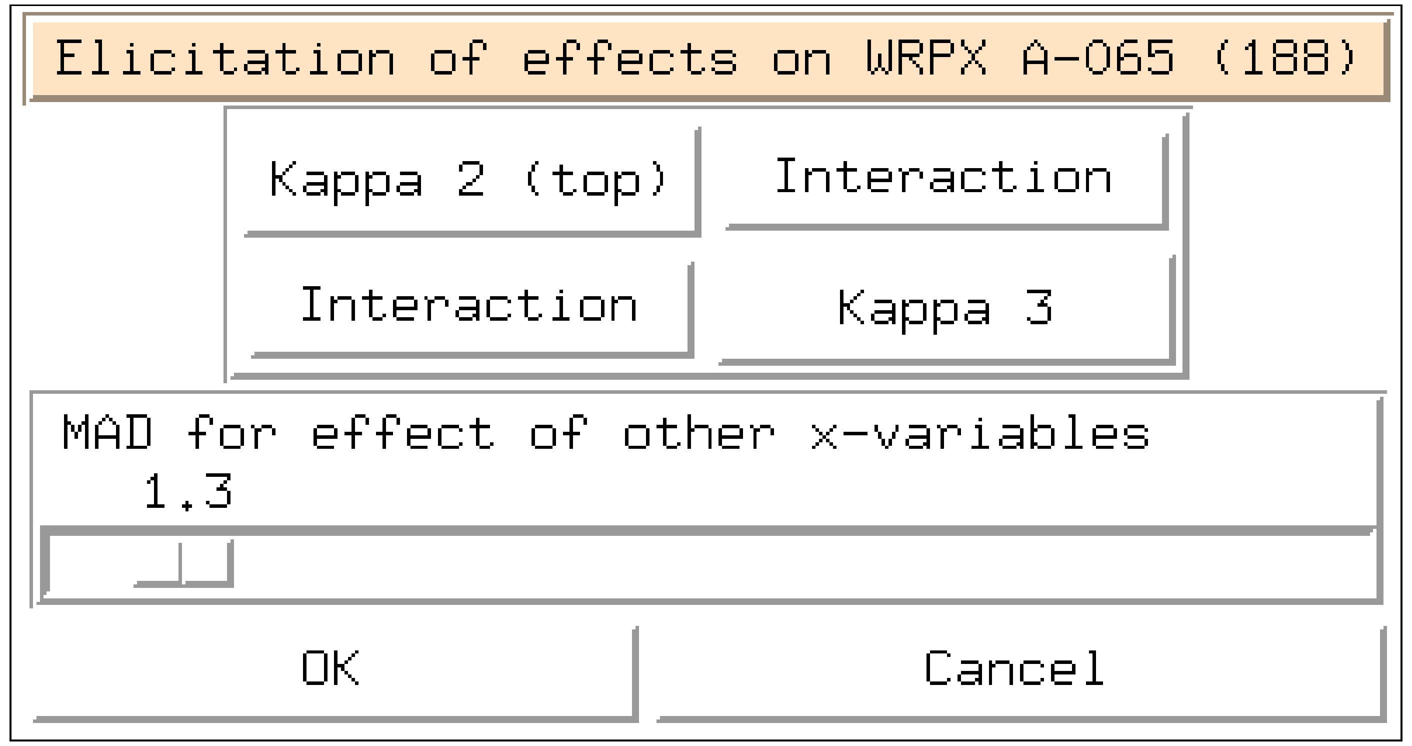


Construction of Groups and Choice of Active Variables

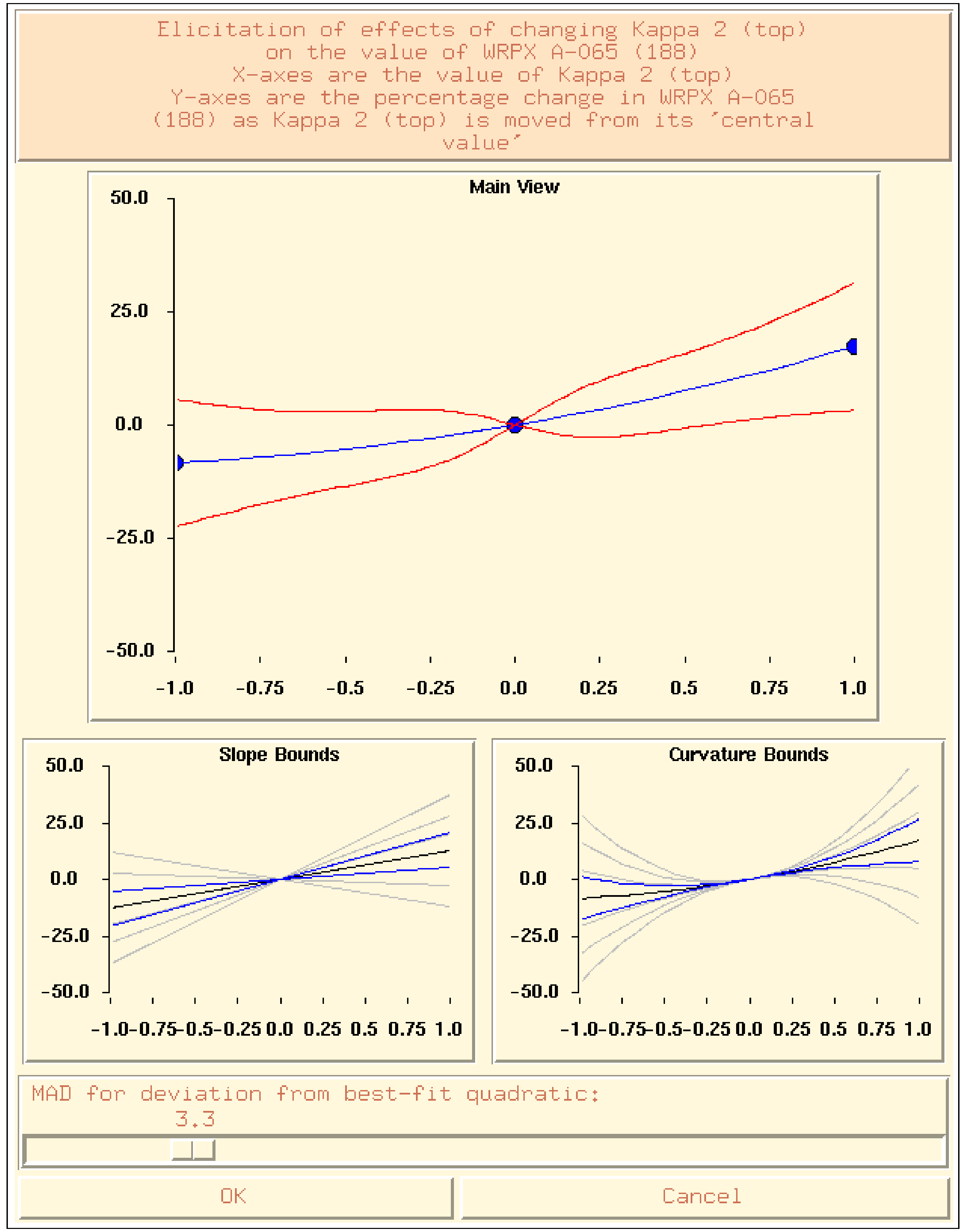




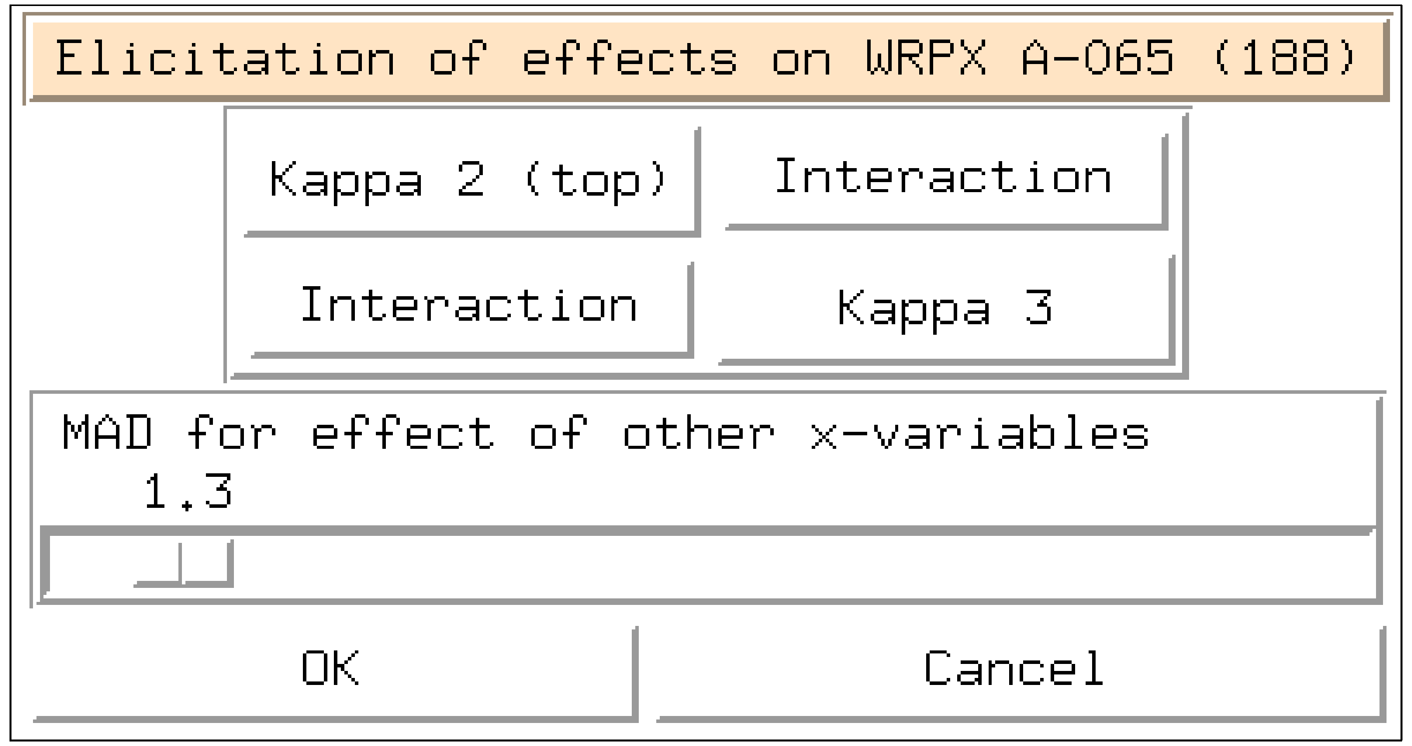
Elicitation for the Master Variable in a Group



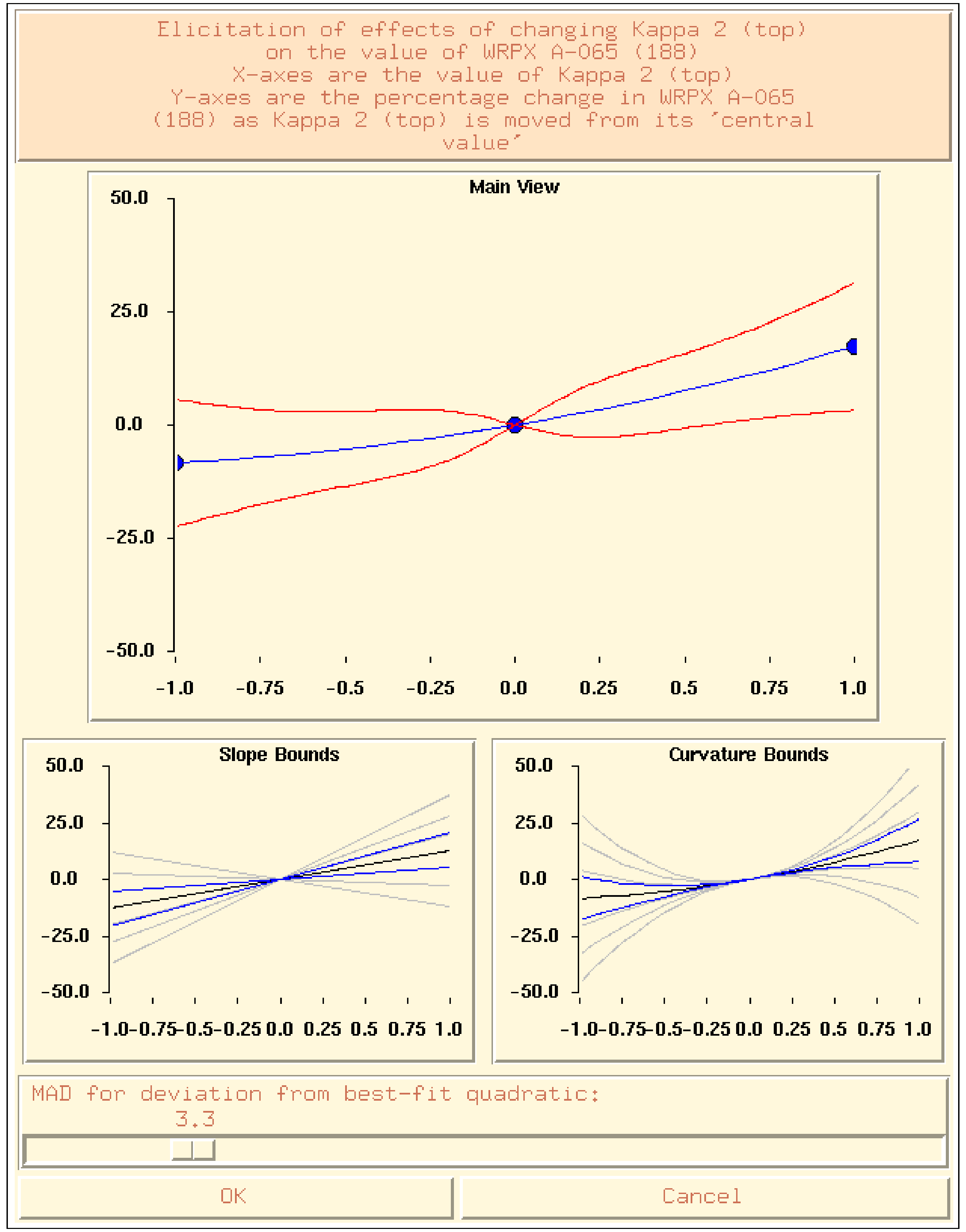
Eliciting Effects of a Single x



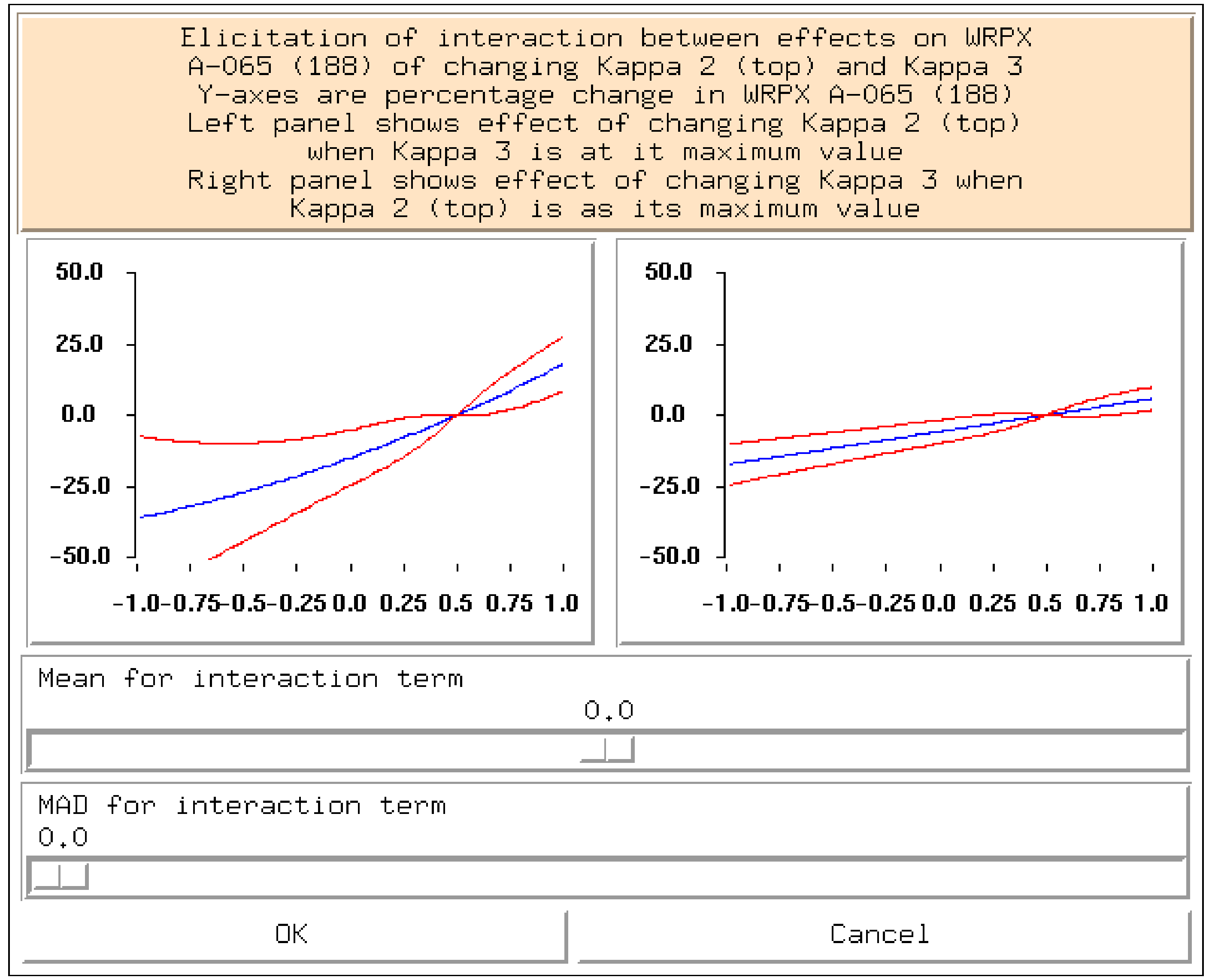
Elicitation for the Master Variable in a Group



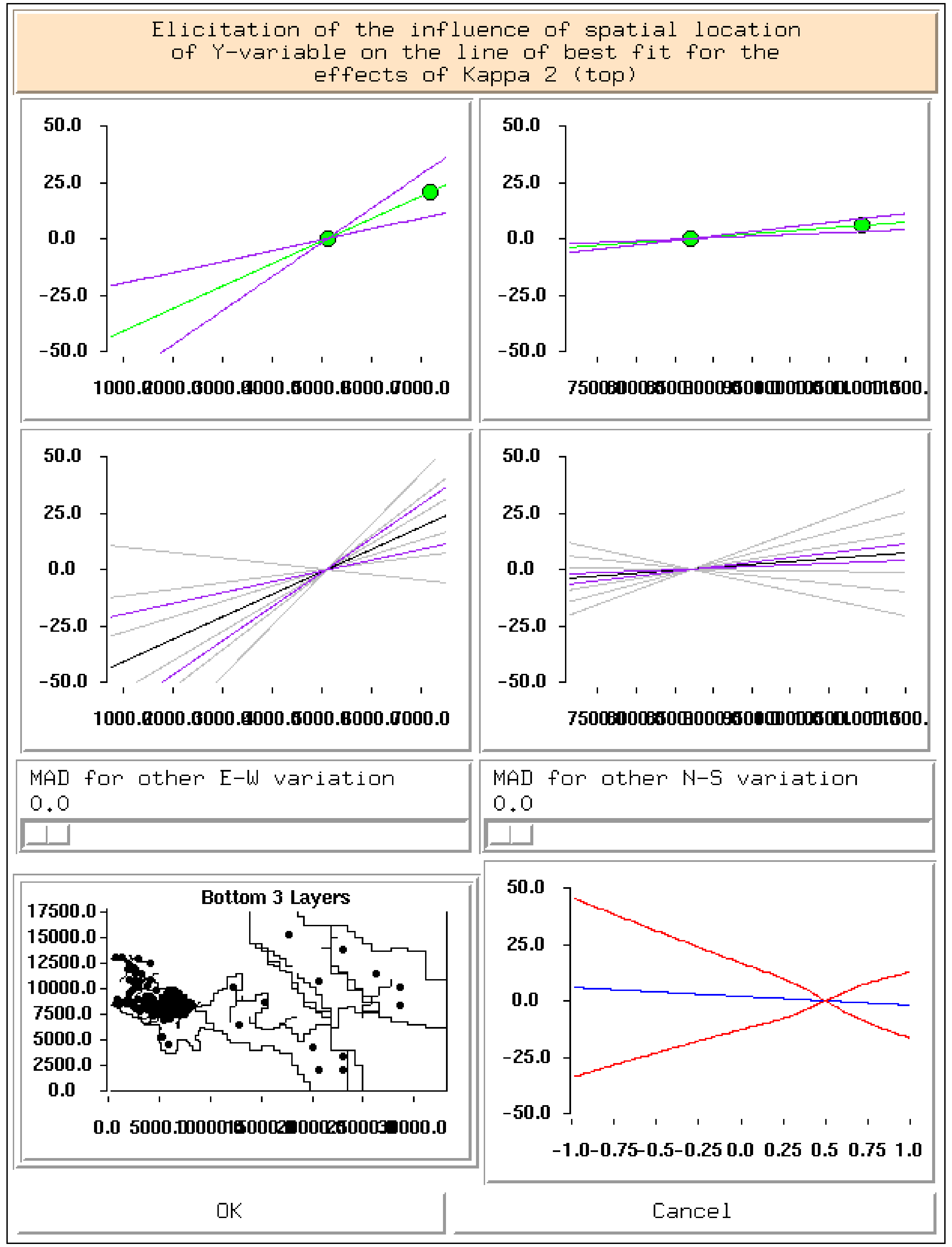
Eliciting Effects of a Single x



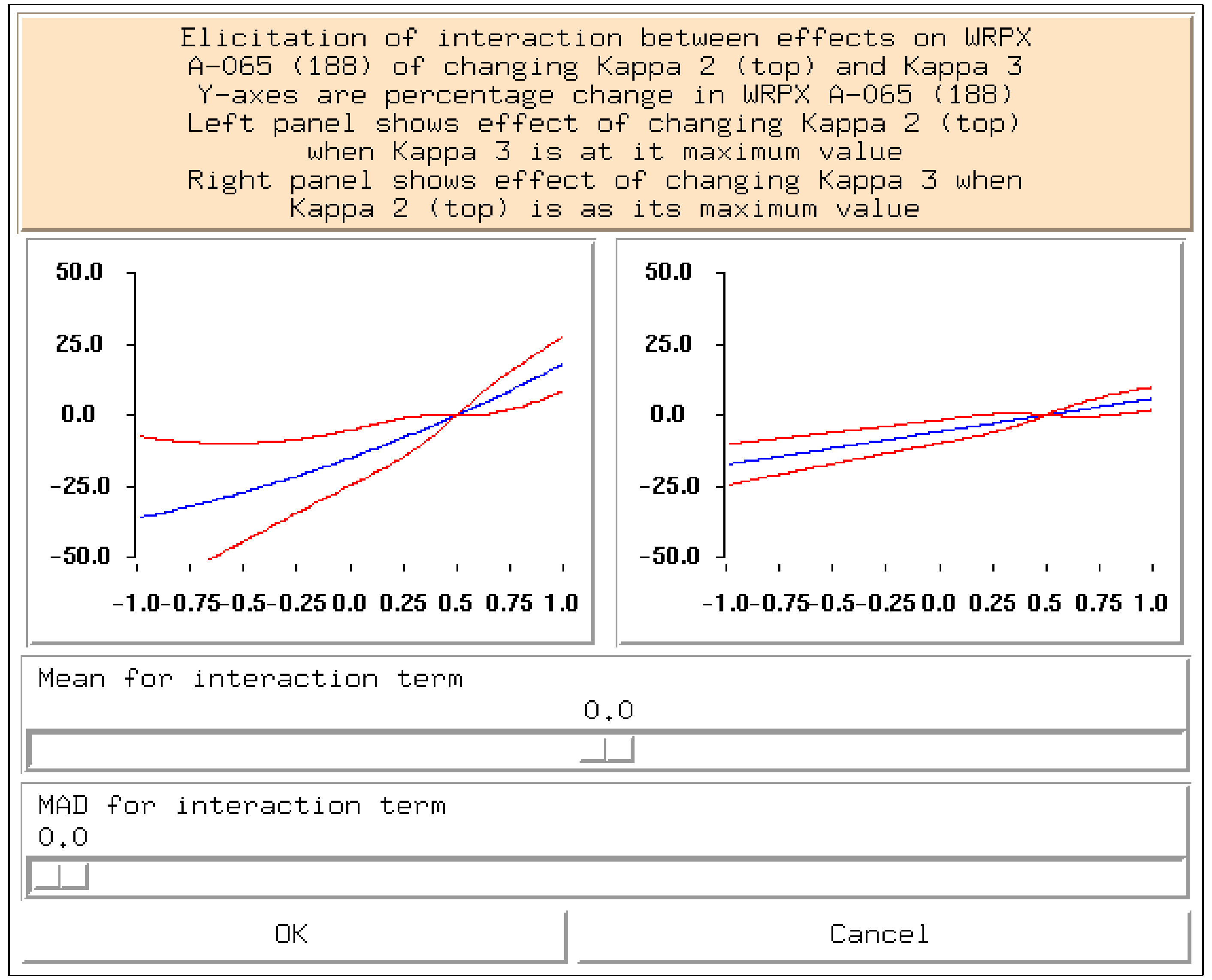
Eliciting Interactions Between x Variables



Eliciting Spatial Changes in the Effects of a Single x



Eliciting Interactions Between x Variables



Eliciting Spatial Changes in the Effects of a Single x