

Robust combination of the Morris and Sobol methods in complex multidimensional models

Dorleta Garcia^{a,*}, Inmaculada Arostegui^b, Raúl Pallezo^a

^aAZTI, Txatxarramendi ugarte z/g, 48395 Sukarrieta, Spain

^bEuskal Herriko Unibertsitatea UPV/EHU, Basque Country, Spain & Basque Center for Applied Mathematics - BCAM

Abstract

Conducting global sensitivity analysis using variance decomposition methods in complex simulation models with many input factors is usually unaffordable. An alternative is to first apply a screening method to reduce the number of input factors and then apply a variance decomposition method to the reduced model. However, usually selection of input factors is not done robustly and convergence of the screening method is not ensured.

We propose two new criteria, a criterion that mimics the visual selection of the input factors and a convergence criterion. In the application of the criteria to a complex model, the Morris screening method has needed 200 trajectories to converge and the visual criterion has outperformed other existing criteria. Our proposal ensures a robust combination of the Morris and the Sobol methods that provides an objective and automatic method to select the most important input factors with a feasible

*Corresponding author: Tel.: +34 667 174 403; fax: + 34 94 687 00 06
Email address: dgarcia@azti.es (Dorleta Garcia)

computing load to achieve convergence.

Keywords: convergence criterion, global sensitivity analysis, Morris elementary effect method, selection criterion, Sobol variance decomposition method

1. Introduction

1

Simulation models are useful tools to provide a better understanding of the environmental systems. One of the big issues when using simulation models is their validation, that is, to ensure the model is good enough to meet its intended purpose (Rykiel, 1996; Schmolke et al., 2010). Balci (1997) presented a long list of techniques for the validation of simulation models divided in four groups: *Informal* techniques which rely on human reasoning and lack mathematical formalism, *Static* techniques which are concerned with accuracy assessment on the basis of characteristics of the static model design and source code, *Dynamic* techniques which evaluate the model based on its execution behavior and *Formal* techniques which

¹List of abbreviations used throughout the manuscript:

- AEE: absolute elementary effect
- CV: coefficient of variation
- GSA: global sensitivity analysis
- SSB: spawning stock biomass
- TAC: total allowable catch

are based on the mathematical proof and correctness. Global sensitivity analysis (GSA) which has been proposed by several authors as a key ingredient in the validation process of simulation models (Saltelli et al., 2000; Rabitz, 1989) is in the third group, the dynamic techniques.

Variance based GSA examines the relation between the variance of the output of the simulation models and the variance of their input factors (Saltelli et al., 2008). Several methods for performing GSA exist, from simple scatterplots to the more complex Sobol variance decomposition method (the Sobol method, Sobol (1993)) (see Pianosi et al. (2016), Borgonovo and Plischke (2016) or Norton (2015) for recent reviews on existing methods and practices). The Sobol method is frequently considered the reference method for variance based GSA (Yang, 2011; Confalonieri et al., 2010; Sarrazin et al., 2016; Homma and Saltelli, 1996). The method can be used to rank the input factors according to their effect on the results and to estimate each input factor's contribution to the output variance. Two of the main drawbacks of this method are its high computational cost and its inability to represent the outputs' uncertainties correctly if the model output is highly skewed (Borgonovo, 2011; Pianosi and Wagener, 2015).

The computational cost of applying the Sobol method on highly non-linear simulation models with many input factors could be unaffordable. In those cases, a frequently used alternative is to combine the Sobol method with the Morris elementary effects method (the Morris method) (Morris, 1991; Campolongo et al., 2007). In this framework, the Morris method is used to select the input factors to be considered later on in the Sobol

method. The Morris method consists of calculating the elementary effect for each input factor on each output variable. Afterwards, the most important input factors are selected visually by identifying the set of input factors that are distinguished from the others because of their high mean absolute elementary effect value. This selection is feasible when the number of output variables is small; however, when this number is high, it can be inaccurate and biased. Furthermore, to assess the convergence or to calculate the confidence intervals using bootstrapping, the selection procedure should be automated.

In the literature a pre-specified number of input factors, the same for all the output variables, is often used to select the most important input factors in each output variable (Hussein et al., 2011; Morris et al., 2014; DeJonge et al., 2012). Campolongo et al. (2007) proposed using Savage scores (Savage, 1956) in the ranking of each output variable and ordering the input factors according to the sum of their scores. However, we have not found any other example using this method in the literature. Both criteria would lead to the selection of unimportant input factors if, for example, for an output variable the variance is explained by only a few number of input factors. The fixed number of factors criterion would select the agreed number of input factors even if some of them are unimportant and the Savage criterion would assign a high score to all the input factors in the top of the ranking even if they have low relevance. Furthermore, the criterion based on Savage scores penalizes the input factors that are important in only one output variable in favor of those that are important in several output variables, even if they are correlated.

Sarrazin et al. (2016) proposed three criteria to assess the convergence of the Morris and Sobol methods. Nevertheless, none of these criteria was designed to ensure the convergence of the Morris method when the objective is to select the most important input factors, being the number of selected input factors equal to a pre-specified number. But this objective arises naturally when the goal is to combine the Morris and Sobol methods to reduce the computational cost of the analysis. In this case, the criteria defined by Sarrazin et al. (2016) could lead to a computational surcharge .

To overcome these problems, new selection and convergence criteria are defined here for the Morris method. The selection criterion, referred here as the calibrated visual criterion, provides a systematic basis for the screening process. In turn, the convergence criterion ensures that the procedure has converged to the group of the most important input factors, which are those that will be proposed to enter into the Sobol method.

The objective of this study is to define a procedure for robustly selecting the input factors that will be considered in the Sobol method after the Morris method is applied. We illustrate the approach using a complex implementation of the FLBEIA bio-economic fisheries simulation model (Garcia et al., 2017b), where the number of input factors is 133. The performance of the calibrated visual criterion is compared with the performance of the criterion that selects a fixed number of factors and the criterion based on Savage scores (from now on the Savage criterion). The reference for the evaluation are the rankings obtained for each output variable using the Sobol method and the ranking for multivariate output models obtained with the method by Lamboni et al. (2011). The convergence of

the Sobol method is assessed using the criterion defined by Sarrazin et al. (2016).

2. Material and Methods

Several methods to perform sensitivity analysis (Saltelli et al., 2008; Norton, 2015) are available. For highly nonlinear and computationally costly models the combination of the Morris and the Sobol methods is recommended (Campolongo et al., 2007); the first to identify the most important input factors at low computational cost and the second to obtain a detailed decomposition of the output variance as a function of the input factors identified by the Morris method.

2.1. *The Morris elementary effects method*

Morris introduced the elementary effects method in 1991 (Morris, 1991) and other authors developed it further (Campolongo et al., 2007, 2011; Ruano et al., 2012). It is an effective mean of identifying important input factors at a lower computational cost than the Sobol method (Saltelli et al., 2008). Campolongo et al. (2007) improved the method's convergence through more efficient sampling of the input space. Furthermore, they developed an expression that allows grouping of input factors and their treatment as if they were a single input factor, with the subsequent reduction in computational cost. Extension of the methods of Campolongo et al. (2007) and criticism of their examples appeared in Norton (2009).

The method consists of evaluating the simulation model, φ , along a set of trajectories, \mathbb{P} , defined in the unit hypercube, $\omega = [0, 1]^K$, where K

corresponds with the number of input factors. When the existence domain of the model is different to the unit hypercube, the trajectories are transformed into the model's original domain, Ω , using a transformation function. The absolute elementary effect (AEE) is calculated for each input factor X_k , for $k \in \{1, \dots, K\}$ and for each trajectory in \mathbb{P} . For simplicity of notation, we will omit the k subscript for the input factor whenever it is not necessary in the context. Therefore, the AEE for input factor X_k and trajectory p is defined as:

$$\text{AEE}_{p, X_k}(\mathbf{X}) = \frac{|\varphi(\mathbf{X}) - \varphi(\mathbf{X}')|}{\Delta} \quad (1)$$

where φ denotes the simulation model, $\varphi(\mathbf{X}) = \mathbf{Y}$ where $\mathbf{Y} = (Y_1, \dots, Y_J)$ represents the output of the model, J denotes the number of output variables, \mathbf{X} and \mathbf{X}' are two consecutive points in the trajectory p that differ only in the value of X_k and Δ is the width of the subintervals in the Morris method. Finally, the AEE of the input factor X_k , AEE_{X_k} , is equal to the mean of the AEEs along all the trajectories:

$$\text{AEE}_{X_k} = \frac{\sum_{p \in \mathbb{P}} \text{AEE}_{p, X_k}}{R} \quad : \quad k \in \{1, \dots, K\} \quad (2)$$

where R denotes the cardinality of \mathbb{P} . The AEE-s are calculated for each output variable. Hence, for each input factor X_k there is a set of absolute elementary effects $\{\text{AEE}_{X_k, Y_j}\}_{j=1}^{j=J}$, where j is the subscript for the output variable. For simplicity of notation, we will omit the j subscript for the output variable whenever it is not necessary in the context.

The following subsections present the calibrated visual criterion to select the most important input factors and the convergence criterion for the

Morris method.

2.1.1. The calibrated visual criterion

First, we define three selection criteria that jointly provide mathematical sense to the criterion used in the visual selection. To give a closed expression for the three criteria, for each output variable Y , we order the input factors according to their AEE value, i.e., $AEE_{X_1,Y} \leq AEE_{X_2,Y} \leq \dots \leq AEE_{X_K,Y}$ and define \mathbb{F} as the set of all the input factors.

1. *Fixed number of factors:* The selected input factors are those that verify that their AEE are among the δ_F input factors with the highest AEE for at least one output variable Y_{j_0} . The set of selected input factors is denoted as \mathbb{F}_F and it is defined as,

$$\mathbb{F}_F = \left\{ X \in \mathbb{F} : \exists j_0 \in \{1, \dots, J\} \text{ s.t. } AEE_{X, Y_{j_0}} > AEE_{X_{K-\delta_F}, Y_{j_0}} \right\} \quad (3)$$

2. *Factors with high AEE value:* The selected input factors are those that verify that their AEE is higher than a proportion, δ_H , of the maximum value of all the AEE-s for at least one output variable Y_{j_0} . The set of selected input factors is denoted as \mathbb{F}_H and it is defined as,

$$\mathbb{F}_H = \left\{ X \in \mathbb{F} : \exists j_0 \in \{1, \dots, J\} \text{ s.t. } AEE_{X, Y_{j_0}} \geq \delta_H \cdot \max\{AEE_{X_k, Y_{j_0}}\}_{k=1}^K \right\} \quad (4)$$

3. *Factors distinguished from the others:* The selected input factors are those that verify that the difference between all the consecutive AEE-s is higher than a proportion δ_D , for all the AEE-s with a higher AEE

than the input factor itself, for at least one output variable Y_{j_0} . The set of selected input factors is denoted by \mathbb{F}_D and it is defined as:

$$\mathbb{F}_D = \left\{ X \in \mathbb{F} : \exists j_0 \in \{1, \dots, J\} \text{ s.t. } \frac{\text{AEE}_{X_k, Y_{j_0}} - \text{AEE}_{X_{k-1}, Y_{j_0}}}{\text{AEE}_{X_k, Y_{j_0}}} \geq \delta_D, \forall X_k : \text{AEE}_{X_k, Y_{j_0}} > \text{AEE}_{X, Y_{j_0}} \right\} \quad (5)$$

Then, given \mathbb{P} a set of trajectories along ω and $K_{EE} < K$ the number of input factors we intend to enter into the Sobol method, the calibrated visual criterion is defined as the weighting of the three criteria defined above and it is applied as follows.

1. Evaluate the model at the points that form the trajectories in \mathbb{P} and calculate the $\{\text{AEE}_{X_k, Y_j}\}_{k=1}^K$ for all $j \in \{1, \dots, J\}$ using equation 2.
2. Find the parameters δ_F , δ_H , and δ_D that result in the selection of K_{EE} input factors. With the *fixed-number of factors* criterion, it may be impossible to select exactly K_{EE} input factors, in which case δ_F is selected as the minimum number of input factors that results in selecting a total number of input factors equal or bigger than K_{EE} .
3. To support calibration of the selection criterion, conduct a visual selection of the input factors. A set of input factors is selected for each output variable and the resulting sets are then merged in a single set \mathbb{F}_V . The selection is done in such a way that the cardinality of \mathbb{F}_V is equal to K_{EE} .
4. Apply the weighted criterion for the 3 previously defined criteria using different combination of weights. Firstly, define a three di-

mensional set of values that provide a good coverage of the unit hypercube. Secondly, for each triplet in the set of weights and each output variable, the number of input factors selected is equal to the weighted mean of those selected with each of the three criteria. Finally, once the number of input factors to be selected for each triplet is decided, the ones with the highest AEE are selected. Then, the set of input factors that corresponds to each triplet of weights, \mathbb{F}_W , is formed by the union of the sets of input factors selected for each output variable.

5. For each triplet compare the corresponding set of input factors calculated in the previous step, \mathbb{F}_W , with \mathbb{F}_V . Then, identify the weights, w_F, w_H and w_D that produce the largest intersection between both sets and among those select the triplet that produces the smallest cardinal of \mathbb{F}_W .

Thus, we obtain a procedure that uses the same criterion for the selection of input factors in all the output variables. Furthermore, the input factors selected with this procedure highly agree with the visually selected ones.

2.1.2. Convergence criterion

We consider that the Morris method has converged when the input factors identified as the most important do not change when the cardinal of \mathbb{P} is increased. We assess convergence using bootstrapping and the selection criterion defined previously.

First, we generate randomly a sufficiently large set of trajectories, \mathbb{P} ,

with cardinal R . Then, using the method in Campolongo et al. (2007) we find the set of trajectories \mathbb{P}_r for different values of r such that $r < R$. In particular, for each i and l such that $r_i < r_l$, once AEE-s are calculated for \mathbb{P}_{r_i} , we need only to evaluate the model in the trajectories that are not included in \mathbb{P}_{r_i} in order to calculate AEE-s for \mathbb{P}_{r_l} .

For each r , we perform the bootstrapping in three steps using N_{boot} iterations:

1. Apply the calibrated visual criterion to \mathbb{P}_r to obtain the weights, w_F , w_H , w_D as proposed for the calibrated visual criterion.
2. Sample with replacement r trajectories from the original set \mathbb{P}_r .
3. Find the value of the parameters δ_F , δ_H , and δ_D as proposed for the calibrated visual criterion.
4. Apply the calibrated visual criterion to that sample using the set of parameters $\{w_F, w_H, w_D, \delta_F, \delta_H, \delta_D\}$ obtained in previous steps.
5. Repeat the process N_{boot} times.

To assess convergence, we define the indicator m_X^r for each r and each input factor X :

$$m_X^r = \sum_{i=1}^{N_{\text{boot}}} \pi_X^r(i) \quad (6)$$

where π_X^r is equal to 1 if input factor X has been selected in iteration i , and 0 otherwise. If an input factor is selected in all the bootstrap iterations, i.e., if $m_X^r = N_{\text{boot}}$, the input factor is among the most relevant ones. Therefore, to identify the K_{EE} most important input factors, it would be

sufficient to increase the number of trajectories r until K_{EE} input factors are selected in all the bootstrap iterations.

However, this condition could be very demanding, and therefore, the criterion can be relaxed using a proportion α of N_{boot} . We define \mathbb{F}_r as the set of input factors selected in, at least, $\alpha \cdot N_{boot}$ bootstrap iterations when r trajectories are used:

$$\mathbb{F}_r = \{X \in \mathbb{F} : m_X^r \geq \alpha \cdot N_{boot}\} \quad (7)$$

If K_r is the cardinality of \mathbb{F}_r , K_r increases with r and we consider that the process has converged when $\exists r_0 \leq R$ such that:

$$K_{r_0} = K_{r_0+1} = \dots = K_{r_{\max}} \quad (8)$$

In general $K_{r_{\max}}$ is lower than K_{EE} because the number of input factors selected in each bootstrap iteration are constrained to result in the selection of K_{EE} input factors. Hence, in general, those selected in $\alpha \cdot N_{boot}$ bootstrap iterations will be equal or lower than K_{EE} .

When convergence has been achieved for the number of input factors to be selected, we define three criteria to select the input factors to be considered when applying the Sobol method, \mathbb{F}_M .

The set of input factors selected with the maximum r , r_{\max} , used in the application of the Morris method:

$$\mathbb{F}_M = \mathbb{F}_{r_{\max}} \quad (9)$$

The union:

$$\mathbb{F}_M = \bigcup_{r=r_0}^{r_{\max}} \mathbb{F}_r \quad (10)$$

The intersection:

$$\mathbb{F}_M = \bigcap_{r=r_0}^{r_{\max}} \mathbb{F}_r \quad (11)$$

The three criteria yield a different number of selected input factors, because in the distribution tail of the AEEs some input factors go in and out of \mathbb{F}_r . In terms of selecting a smaller number of input factors, the most restrictive option is the third, whereas the second is the most conservative, and the first is intermediate. As a general procedure, we can examine the degree of difference between the three options in terms of the set of selected input factors.

Figure 1 shows the application of the whole proposal including the two criteria, the calibrated visual criterion for selection and the bootstrap for convergence.

2.2. Sobol variance decomposition method

The Sobol method consists of the decomposition of the output variance as a function of the variance of the conditional expectations of the output on any combination of input factors (Sobol, 1993). Homma and Saltelli (1996) proposed summarizing the contribution of the input factors to the output variance using two sensitivity indices: *first-order* and *total-effects*. The first is equal to the ratio between the variance of the conditional expectation of the model output on k -th input factor and the total variance

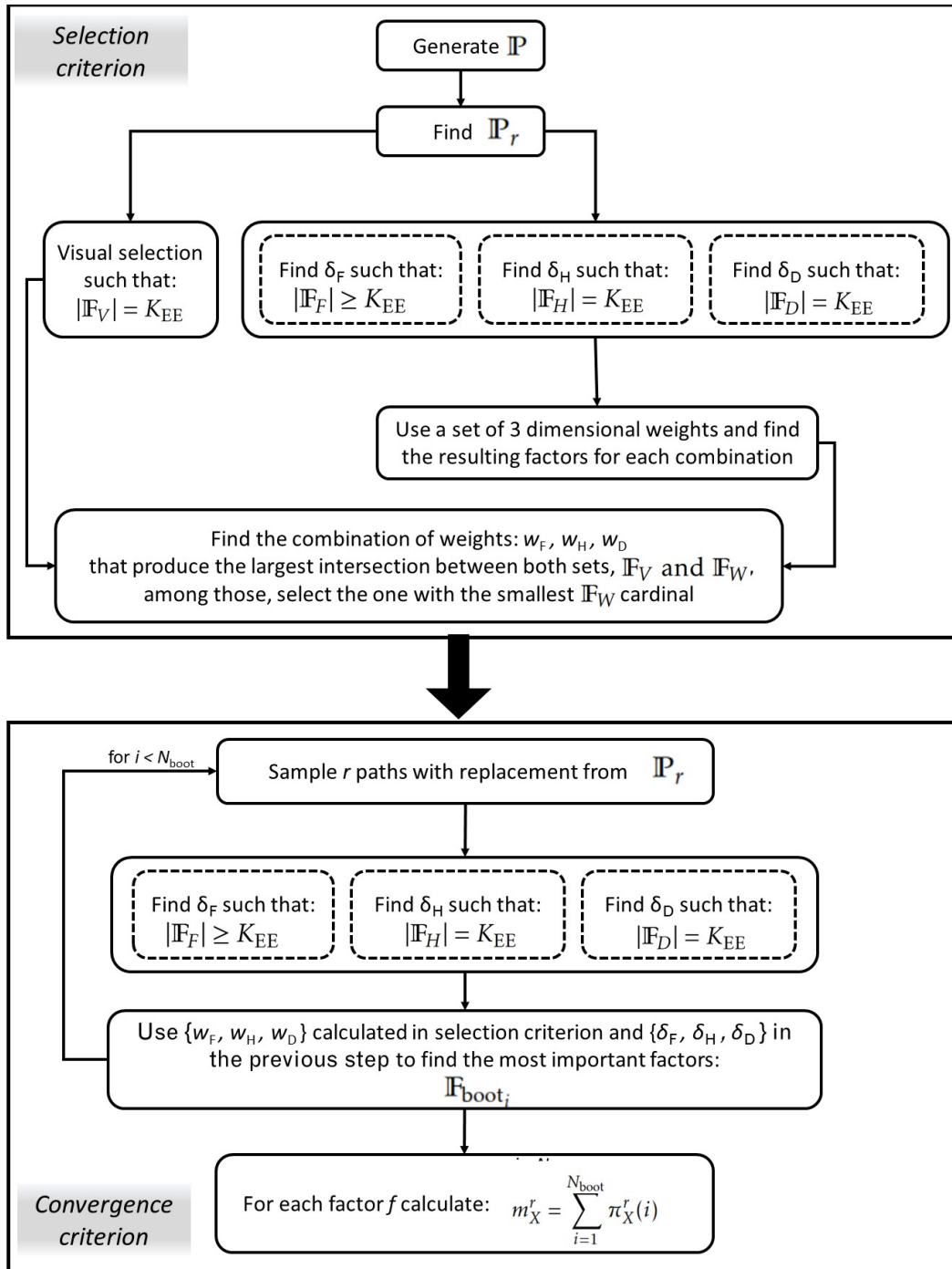


Figure 1: Steps for the application of the selection and the convergence criteria given \mathbb{P} a set of trajectories and r the number of trajectories to use in the analysis.

of the model output, mathematically:

$$S_k = \frac{V(E(Y|X_k))}{V(Y)} \quad (12)$$

where X_k denotes the k -th input factor, $Y = \varphi(\mathbf{X})$ is the unidimensional output of the simulation model represented by φ and $\mathbf{X} = (X_1, \dots, X_K)$ represents the model input. This index represents the contribution of the k -th input factor to the output variance in isolation.

In turn, the *total-effect* is equal to the expected value of the conditional variance of the model output on all the input factors but one, the k -th input factor, denoted here as $\mathbf{X}_{\sim k}$. It represents the contribution to the variance of the k -th input factor alone and in combination with the remaining input factors. Mathematically it is written as:

$$S_{T_k} = \frac{E(V(Y|\mathbf{X}_{\sim k}))}{V(Y)} \quad (13)$$

In simple cases, the sensitivity indices can be calculated analytically. However, in most cases the models are too complex to allow the derivation of analytical expressions for equations 12 and 13. Saltelli et al. (2010) compared different approaches for calculating the Sobol sensitivity indices using Monte Carlo simulations. Here, we use the approach that was identified as the best in terms of the convergence rate.

First, two independent matrices of dimension $N \times M$ are constructed, A and B , the so-called *sample* and *re-sample* matrices, where N and M are the number of base simulations and input factors of the model, respectively. The input factors can be multivariate, and therefore, M can be larger than the number of input factors in the model, K . When the input factors are

divided in groups, instead of considering every input factor alone, the elements in the Sobol decomposition that include this input factor represent the contribution to the variance of all the input factors in the group as a whole, in isolation in the case of *first-order* index, and in combination with other sets of input factors, in the case of the rest of the elements in the decomposition of variance. Hence, the input factors should be grouped sensibly to obtain meaningful results.

Second, additional K matrices, $\{C_k\}_{k \in 1, \dots, K}$, are constructed from the A and B matrices. Each C_k matrix is equal to A , except in the columns that correspond to the k -th input factor, which are taken from matrix B . Finally, the model is applied to each of the rows of A , B , and $\{C_k\}_{k \in 1, \dots, K}$ matrices. The numerator in equation 12 is then approximated by:

$$V(E(Y|X_k)) = \frac{1}{N} \sum_{i=1}^N \varphi(B_{i.}) \cdot (\varphi(C_{k,i.}) - \varphi(A_{i.})) \quad (14)$$

where $A_{i.}$, $B_{i.}$ and $C_{k,i.}$ denote the i -th row of matrices A , B , and C_k , respectively. In turn, the numerator in equation 13 is estimated as:

$$E(V(Y|\mathbf{X}_{\sim k})) = \frac{1}{2N} \sum_{i=1}^N (\varphi(C_{k,i.}) - \varphi(A_{i.}))^2 \quad (15)$$

Finally, the total variance $V(Y)$ is approximated by:

$$V(Y) = \frac{1}{N} \sum_{i=1}^N \varphi(A_{i.})^2 - \left(\frac{1}{N} \sum_{i=1}^N \varphi(A_{i.}) \right)^2 \quad (16)$$

The convergence of the estimators can be assessed using the bootstrap confidence intervals' width (Sarrazin et al., 2016).

2.2.1. Multivariate outputs

The generalized sensitivity indices proposed by Lamboni et al. (2011) are the equivalent of the sensitivity indices defined in the previous section but for the overall variance of the output of a model with a multidimensional output. The generalized indices are based on the work of Campbell et al. (2006) who proposed to decompose the multivariate output in an orthogonal system and then apply the sensitivity indices to the most informative components individually. Lamboni et al. (2011) developed further the idea proposed by Campbell et al. (2006) and using principal component analysis as orthogonal decomposition, proved that the *first-order* and *total-effect* indices calculated on the sum of the principal components are to multivariate outputs what the Sobol sensitivity indices are to the univariate one.

2.3. Performance of the selection criterion

Two performance indicators are defined to evaluate the performance of the calibrated visual criterion and other two selection criteria, the selection of a fixed number of factors for each output variable and the criterion based on Savage scores (Campolongo et al., 2007). They are based on the *total-effect* indices calculated on the reduced simulation model obtained introducing variability exclusively in the K_{EE} input factors selected with the Morris method. The first performance indicator, uses the set of Sobol's *total-effect* sensitivity indices for each output variable Y_j , $\mathbf{S}_T^j = \{S_{T_k}^j\}_{k=1}^{K_{EE}}$, to assess the performance of the criterion, where $S_{T_k}^j$ denotes the *total-effect* of the k -th input factor for output variable Y_j . In turn, the second one, the generalized performance indicator, uses the *generalized total-effect* indices

for multivariate output defined by Lamboni et al. (2011), $\mathbb{G}_T = \{G_{T_k}\}_{k=1}^{K_{EE}}$, where G_{T_k} denotes the *generalized total-effect* index of k -th input factor.

To assess the performance of the criterion under different conditions, the performance indicators are calculated for different sets of output variables Γ and different number of input factors in the Morris method. Let us Z denote the number of input factors used in the fixed number of factors criterion to calculate the set of input factors in the Morris method. Then, the performance indicators are calculated as follows:

- The fixed number of factors criterion is applied to the Morris elementary effects selecting the Z input factors with the highest elementary effect value. The resulting number of selected input factors is denoted as $K_{EE,Z}$.
- The calibrated visual criterion is applied using $K_{EE,Z}$ number of input factors as threshold.
- The Savage criterion is applied selecting the $K_{EE,Z}$ input factors with the highest score.
- For a given selection criterion, to calculate the performance indicator for output variable Y , first the corresponding *total-effect* values are assigned to the input factors selected in the application of the criterion, $\{X_1, \dots, X_{K_{EE,Z}}\}$, i.e :

$$\rho_k^j = \begin{cases} 0, & \text{if } X_k \notin \mathbb{F}_M. \\ \frac{S_{T_k}^j}{\sum_{i=1}^{K_{EE}} S_{T_i}^j}, & \text{otherwise.} \end{cases} \quad (17)$$

where $S_{T_k}^j$ corresponds with the *total-effect* of input factor X_k for output variable Y_j . Then, the first performance indicator, Θ , is calculated as the ratio between the sum of all the ρ_k^j over all the input factors selected by the criterion and all the output variables in Γ . The sum is then divided by the number of output variables to place the possible values of the indicator between 0 and 1.

$$\Theta = \frac{1}{|\Gamma|} \sum_{j=1}^{|\Gamma|} \sum_{k=1}^{K_{EE,Z}} \rho_k^j \quad (18)$$

The second performance indicator, the generalized indicator, Θ_G , is calculated similarly but instead of having one *total-effect* index per output variable Y , there is only one *total-effect* index for all the output variables. Hence, ρ depends only on the input factors and in equation 18 the sum along output variables and the division by the number of output variables disappear.

In the comparison of the three criteria the one with the highest Θ is the criterion which produces the best selection of input factors. Values of Θ equal to 1 indicate that the input factors selected by the criterion are the $K_{EE,Z}$ input factors in the top of the ranking, for all the output variables in the case of the first indicator, and for the ranking obtained with the *generalized total-effect* index in the case of generalized one. The procedure is not applied to $Z = 1$ because it implies to select δ_H and δ_D in such a way that only one input factor per output variable is selected, i.e., the three criteria are equivalent.

2.4. Illustrative Example

2.4.1. General description

The approach is illustrated using a complex implementation of FLBEIA (Garcia et al., 2017b), a bio-economic simulation model that is used to describe fishery systems. In FLBEIA the fishery system is divided in two main components, the real system that includes the fish stocks and the fishing fleets, and the management system that is formed by the data collection, the assessment model and the management advice. The main components of FLBEIA are represented in Figure 2. All the variables in the real system are subject to natural variability and the variables observed in the management procedure are subject to epistemic uncertainty.

The model has been applied to the demersal fishery operating around the Iberian Peninsula in southern Europe. This fishery comprises seven fleets, the activity of which is divided into segments called metiers. The model includes explicitly the stocks caught by the fleets for which absolute estimates of abundance are available: hake, horse mackerel, four spot megrim, megrim, and monkfish. Furthermore, the model includes three widely distributed stocks, western horse mackerel, mackerel, and blue whiting, because of their economic relevance. However, as the catch extracted is a marginal part of the total catch of these stocks, the impact on their biomass is minor and therefore, it has been assumed constant along the simulation. The remaining stocks have been aggregated, at metier level, in one total stock called OTH. As no abundance estimate for any of these stocks exists, it has been assumed that the catch is a function of the metier's effort and independent of biomass.

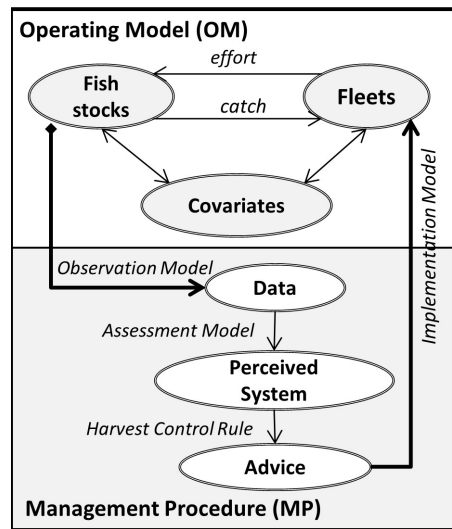


Figure 2: Conceptual diagram of the management strategy evaluation approach. The operating model represents the fishery system and the management procedure reproduces the actual management process step by step. The ellipses correspond to the components of the FLBEIA model. The diagram has been taken from Garcia et al. (2017b)

A brief description of the submodels used to describe the processes that constitute this specific implementation of FLBEIA is given in Table 1. A detailed description of the case study appears in Garcia et al. (2017a).

Table 1: Models used for each stock and fleet in each model component.

| Component | Stock/Fleet | Models Used | |
|--------------------|----------------------|---|--|
| Operating Model | Biological | Hake H.Mackerel 4 Spot Megrim Megrim Monkfish | Exponential Survival & Ricker Recruitment Model Exponential Survival & Segment Regresion Recruitment Model |
| | | Western H. Mackerel Mackerel Blue Whiting | Age Structured Fixed Population |
| | | OTH | No population dynamics |
| | Fleet | Trawlers Gillnetters Longliners | Multi-metier fleets. Effort share given as input data Total effort based on the quota share of all the stocks Entry-Exit model |
| | | P. Trawlers P. Polivalent S. Purse Seiners P.Purse Seiners | Single Metier fleets Total effort based on the quota share of the target stock No economics considered. |
| | Management Procedure | Observation | Hake H.Mackerel 4 Spot Megrim Megrim Monkfish |
| Rest of the stocks | | | Not observed |
| Assess. | | All the stocks | No assessment model |
| Advice | | Hake H.Mackerel 4 Spot M. Megrim Monkfish | The harvest control rule (HCR) used by International Council for the Exploration of the Sea (ICES) in the framework of MSY. |
| | | Western H. Mac. Mackerel Blue Whiting | The historical TAC with uncertainty |

2.4.2. Uncertainty Conditioning

Table A.5 in Appendix A shows a description of the $K = 133$ input factors of the model. Some of them are single input factors and others correspond with a set of input factors introduced in the model as a group

(Cariboni and Campolongo, 2004). As a general rule, a uniform distribution has been used to simulate uncertainty in the input factors. The exceptions are maturity and retention curves, effort share along metiers, and aging error. The values of maturity and retention curves have been simulated using a beta distribution and effort share and aging error using a Dirichlet distribution. The parameters of the distributions have been obtained constraining the mean to the value in Garcia et al. (2017a) and the coefficient of variation (CV) to a 30%. In the case of multiplicative observation errors, that are not included in Garcia et al. (2017a), a mean equal to one has been used, i.e, the errors are unbiased. The aging error has been modeled using a square matrix in which elements a_{il} describe the probability of assigning age i to a fish of age l , and corresponds to the expected value of the Dirichlet distribution. The matrix is the "noise-only, unbiased" matrix in Reeves (2003). Following recommendations in Saltelli et al. (2010), we have sampled the unit hypercube using the Sobol pseudo-random sequences (Sobol, 1967) to accelerate convergence. For univariate input factors, the values have been transformed from the unit hypercube to the original space Ω using inverse transformation method. The conditioning and transformation of effort share and aging error has been done using the procedure proposed in Devroye (1986) for Dirichlet distribution.

2.4.3. *Output variables*

The output of the simulation model has been summarized using five variables per stock and four variables per fleet, which results in $J = 37$ output variables per year. The stock variables are the spawning stock biomass

(SSB) and recruitment, which are related to the stock abundance, fishing mortality, and catch, which are representative of their exploitation level, and the total allowable catch (TAC), which is the output of the decision-making process. The fleets' performance has been summarized using effort, profits, gross value added, and number of vessels. Effort represents the fleets' activity, the profits represent their economic performance, the gross value added is a measure of the goods produced by the fishing activity and the number of vessels shows the variation in the capital.

3. Results

3.1. *Morris Elementary Effects Method*

First, we have generated a set of $R = 1000$ trajectories, \mathbb{P} , along ω . Then, for $r = 25, 50, 100, 150, 200, 250, 300$, we have applied the procedure described in Campolongo et al. (2007) to find \mathbb{P}_r .

In terms of illustration, the calibrated visual criterion has been applied with the objective of reducing the number of input factors to the half, i.e., $K_{EE} = 67$. However, any other objective would be also valid. A set of weights covering the unit cube with intervals of 0.01 width has been used for the weighting procedure. Several weight combinations produce the best match with the visual selection, and therefore, the combination that minimizes the euclidean distance to $(1/3, 1/3, 1/3)$ has been chosen. For $r = 300$, the greatest weight, 0.53, has been assigned to the fixed number of factors criterion, 0.35 to the high AEE value criterion and 0.12 to the difference in AEE criterion.

The method's convergence has been assessed using a bootstrap with

$N_{\text{boot}} = 500$ iterations. The number of input factors selected in the 500 iterations increases quickly with the number of trajectories. With $r = 25$ trajectories, only 19 input factors have been selected in all the iterations and with $r = 300$ this number has increased to 50 (Figure 3). When the criterion is relaxed to 95% of the iterations, for $r = 25$, 42 input factors are selected and then the number of input factors increases steadily and becomes stable at 55 input factors for $r \geq 200$. The sets $\mathbb{F}_{200}, \mathbb{F}_{250}$, and \mathbb{F}_{300} differ in one factor. This occurs because the difference in the value of the AEEs of the input factors that are not in the top of the ranking is so small that the ranking in those positions needs more iterations to stabilize. Hence, to be cautious, we have used the union criterion for $r \geq 200$, which results in the selection of the 56 input factors listed in Table 2.

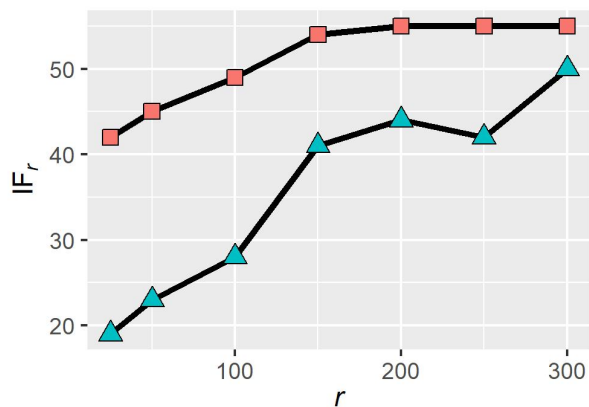


Figure 3: Number of input factors selected in all the bootstrap iterations (triangles) and the number of those selected in 95% of the iterations (squares) as a function of the number of trajectories used in the computation of the elementary effects.

Although the objective is to select 67 input factors, only 56 have been

Table 2: Input factors identified as important by the Morris elementary effects method. SRR = Stock recruitment relationship, Obs. = Observation, HKE = hake, HOM = horse mackerel, LDB = four spot megrim, MEG = megrim, MON = monkfish, PTB = Pair trawlers, DTS = Trawlers.

| Factors | Stock Level | |
|-------------------------|--------------------|--|
| | Stock | |
| Aging Error | all | |
| Maturity | HKE, MON | |
| Natural Mortality | all | |
| Weight | all | |
| Initial Population | HKE, HOM | |
| SRR parameters | all | |
| Uncertainty around SRR | all | |
| Obs. error in abundance | HKE, HOM, MEG, MON | |
| Obs.error in weight | all | |
| TAC | MAC, HO8 | |

| Factor | Fleet Level | |
|---------------------|-------------|--|
| | Fleet | |
| Crewshare | ALL | |
| Effortshare | ALL | |
| FuelCost | DTS | |
| Maximum days at sea | ALL | |
| w1 | DTS | |

| Factor | Fleet-Metier and Stock level | |
|-------------|------------------------------|----------------------|
| | Stock | Fleet-metier |
| Cachability | HKE | PTB metier in DTS SP |
| Cachability | HO8, LDB, MAC, MEG | all |

finally selected. The number of input factors selected with the calibrated visual criterion in each bootstrap iteration varies between 60 and 72, with median equal to 66 and mode equal to 67. The number of input factors is not always equal to 67. Even if the number of input factors selected by the fixed number of factors criterion is 67, the application of the weighted criterion does not guarantee that the number of input factors selected is 67. This happens because whereas the restriction of selecting 67 input factors is applied to the three criteria globally, the weighted criterion is applied at output variable level. Afterwards, in the analysis of convergence only 56

input factors have been selected in 95% of the bootstrap iterations. Therefore, there is a set of more than 10 input factors entering and leaving the group of the most important 67 input factors. That is, the number of input factors in the calibrated visual criterion should be increased to end up with a larger group of selected input factors. However, the difference between the input factors that are not in the top of the ranking is that small that the ranking of those input factors is difficult to stabilize.

In general, for recruitment, spawning stock biomass, TAC, and number of vessels, there is a set of input factors that are differentiated from the rest because of their higher AEE value (see the graphs in the supplementary material S1, S2 and S3). However, for the remaining variables the differentiation is not equally clear. For most of the output variables, the difference between the number of input factors selected visually and those selected with the calibrated visual criterion is equal or lower than one (Table 3). The biggest difference is obtained in the profits of trawlers where visually 12 input factors are selected and with the calibrated visual criterion only 7. When a set of input factors exists that is clearly distinguished from the rest, the visual selection is more precise. However, when the differentiation between sets is unclear, in the case of fishing mortality and output variables relative to hake, for example, the calibrated visual criterion tends to select more input factors. Furthermore, the variability in the number of input factors selected is higher for the visual criterion and the number of input factors selected is lower in general. Although variable by variable some differences exist between the visual and calibrated criteria, as the input factors are aggregated in a single set and the most important

input factors appear at the top of many of the variables, at the overall level the differences are small.

Table 3: Number of input factors selected by each of the selection criteria for each output variable. The column 'Fixed' corresponds with the *fixed number of factors* criterion, 'High' with the *factors with high AEE value* criterion, 'Diff' with the *factors distinguished from the others* criterion, 'Visual' with the number of factors selected visually and 'Calib.' with the *visual calibrated* criterion. (F = fishing mortality, Rec. = Recruitment, Prof. = Profits, Eff. = Effort, nVes. = Number of vessels, HKE = hake, HOM = horse mackerel, LDB = four spot megrim, MEG = megrim, MON = monkfish, DFN = gillnetters, DTS = trawlers and HOK = long-liners).

| | | Fixed | High | Diff. | Vis. | Calib. |
|-------|-----|-------|------|-------|------|--------|
| SSB | HKE | 6 | 5 | 2 | 5 | 5 |
| | HOM | 6 | 2 | 6 | 6 | 5 |
| | LDB | 6 | 5 | 8 | 8 | 6 |
| | MEG | 6 | 5 | 9 | 5 | 6 |
| | MON | 6 | 5 | 8 | 8 | 6 |
| Catch | HKE | 6 | 16 | 5 | 5 | 9 |
| | HOM | 6 | 6 | 11 | 7 | 6 |
| | LDB | 6 | 9 | 1 | 6 | 6 |
| | MEG | 6 | 17 | 9 | 6 | 10 |
| | MON | 6 | 12 | 6 | 5 | 8 |
| F | HKE | 6 | 19 | 5 | 9 | 10 |
| | HOM | 6 | 7 | 1 | 5 | 6 |
| | LDB | 6 | 10 | 6 | 5 | 7 |
| | MEG | 6 | 9 | 6 | 4 | 7 |
| | MON | 6 | 32 | 4 | 14 | 15 |
| Rec. | HKE | 6 | 8 | 8 | 6 | 7 |
| | HOM | 6 | 3 | 6 | 6 | 5 |
| | LDB | 6 | 2 | 2 | 7 | 4 |
| | MEG | 6 | 2 | 2 | 2 | 4 |
| | MON | 6 | 3 | 4 | 3 | 5 |
| TAC | HKE | 6 | 8 | 9 | 6 | 7 |
| | HOM | 6 | 5 | 7 | 7 | 6 |
| | LDB | 6 | 5 | 9 | 5 | 6 |
| | MEG | 6 | 7 | 16 | 7 | 7 |
| | MON | 6 | 10 | 6 | 5 | 7 |
| Prof. | DFN | 6 | 14 | 3 | 8 | 8 |
| | DTS | 6 | 9 | 4 | 12 | 7 |
| | HOK | 6 | 11 | 9 | 8 | 8 |
| Eff. | DFN | 6 | 21 | 8 | 8 | 11 |
| | DTS | 6 | 10 | 6 | 10 | 7 |
| | HOK | 6 | 13 | 9 | 8 | 9 |
| GVA | DFN | 6 | 16 | 1 | 7 | 9 |
| | DTS | 6 | 8 | 6 | 11 | 7 |
| | HOK | 6 | 11 | 3 | 8 | 7 |
| Nves. | DFN | 6 | 1 | 5 | 5 | 4 |
| | DTS | 6 | 1 | 1 | 5 | 4 |
| | HOK | 6 | 1 | 4 | 4 | 4 |

The application of the Morris method results in the selection of most of the biological input factors (24 input factors out of 35, 69%). On the contrary only a few economic input factors have been selected (5 out of

33, 15%). In the observation error category almost half of the input factors have been selected (16 out of 34, 47%) and in the case of technical input factors one third (11 out of 33, 33%). The uncertainty derived from observation errors, identified as important by the morris method in many cases, can be reduced improving the sampling programs and the mathematical models used to estimate the stock status. In this sense, a variance decomposition GSA including those input factors, would provide the basis to carry out a cost benefit analysis of improving the assessment process of these stocks. Although the uncertainty in the rest of the selected input factors cannot be reduced, this analysis highlights the importance of considering uncertainty in these input factors when the performance of management strategies is evaluated in the long-term. For example, natural mortality is often considered constant and has been identified as an important input factor for all the stocks. The uncertainty related with the recruitment process has been classified as important in all the cases, in line with common practice. The TACs of the pelagic stocks, considered non-target stocks for this fishery and included in this analysis as secondary stocks, have been identified as one of the most important input factors, in line with the claims of the fishing sector. Most of the economic input factors have been rejected. One of the reasons could be that the fleet dynamic model used to predict the effort allocation of the fleets does not consider any economic incentive and economic input factors are simply used to transform the fish tons caught into monetary terms. The effort share input factor has been identified as important for all the fleets, stressing the importance of considering fleet dynamic models in this kind of simula-

tion models. Catchability, the input factor that measures the productivity of the fleets, has been selected only in one third of the cases, in fact, the input factor that differed in \mathbb{F}_{250} and \mathbb{F}_{300} is the catchability of pair trawlers on hake.

The AEEs for all the input factors and output variables for $r = 300$ are provided as supplementary material in a Shiny application (<https://aztigps.shinyapps.io/GSAApp/>, password: flbeiaGSA, the code and data to run the application locally can be downloaded from Zenodo <https://zenodo.org/record/3402534> (Garcia, 2019)).

3.2. Sobol Variance Decomposition Method

We have analyzed the convergence of the Sobol sensitivity indices, examining the width of the bootstrap confidence intervals, as proposed by Sarrazin et al. (2016). The width decreases rapidly with the number of base simulations N for $N < 2000$ (Figure 4). For $N = 150$, the width of the confidence interval of the *total-effect* index of all the input factors and output variables in 2020 is greater than 0.5, but for $N = 1500$, 75% of the intervals are already narrower than 0.05. However, the decrease rate slows for $N \geq 2000$ and for $N = 10000$, 4% of the confidence intervals are wider than 0.05 (Figure 4).

In general, most of the variance of the output variables is explained by the interaction between input factors. The number of vessels, the recruitment and the SSB are the variables of which the variance is explained by the smaller number of input factors. On the opposite side, the variance of the output variables related with effort, fishing mortality, effort itself, profits and GVA, is explained by a great number of input factors (Figure

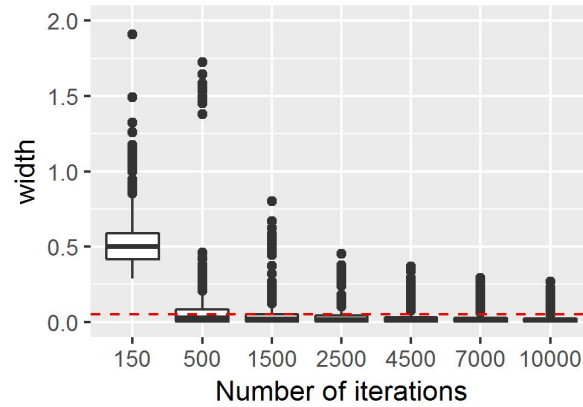


Figure 4: Boxplot of the width of the confidence intervals of the *total-effect* index of the input factors and output variables in 2020. The x-axis correspond with the base sample size N used. The red dashed line indicates the 0.05 threshold for the width.

5).

Once the sensitivity indices have been calculated for $N = 10000$ we have used the method proposed by Lamboni et al. (2011) to calculate the generalized global sensitivity indices using the output variables in 2020. The main result obtained at output variable level is corroborated by the global index: the output variance is largely explained by the interaction between input factors (Figure 6). When the 37 variables are used, thirty input factors are "lower sensitivity" factors (contributing less than 5% to the overall variance (Sarrazin et al., 2016)), i.e., only 26 input factors contribute considerably to the output variance. The *total-effect* of the *generalized sensitivity index* has been used to calculate the performance indicators of the selection criterion. As the index depends on the output variable used, it has been calculated for each set of the output variables.

A complete set of barplots with the *first-order* and *total-effect* indices and their confidence intervals is available in a Shiny application (<https://aztigps.shinyapps.io/GSAApp/>, password: flbeiaGSA, the code and data to run the application locally can be downloaded from Zenodo <https://zenodo.org/record/3402534> (Garcia, 2019)).

3.3. Performance of the selection criterion

The individual and overall level output variables defined in section 2.3 have been calculated for $Z = 2, 3, 4$ and for the three criteria, the calibrated visual criterion, the fixed number criterion and the savage criterion. For $Z > 4$, the number of input factors selected with the calibrated visual criterion is higher than 56. Hence, it makes no sense to calculate the performance indicator because all the input factors selected by the Morris method are selected by the calibrated visual criterion.

Furthermore, we have evaluated the sensitivity of the performance of the calibrated visual criterion to the choice of the output variables. We take three subsets of the output variables, calculate the corresponding *generalized sensitivity indices* and apply the selection criterion using the output variables selected to calculate the performance indicator. In the first set we use all the output variables, i.e., a set with 37 variables. In the second subset with 29 variables, we remove the fishing mortality and the gross value added from the output variables because they are highly correlated with the other variables. In the third subset with 21 variables, besides fishing mortality and gross value added we also eliminate catch and effort. Hence, in the subsets with 21 and 29 variables we remove the output variables that are highly correlated with the rest. Furthermore,

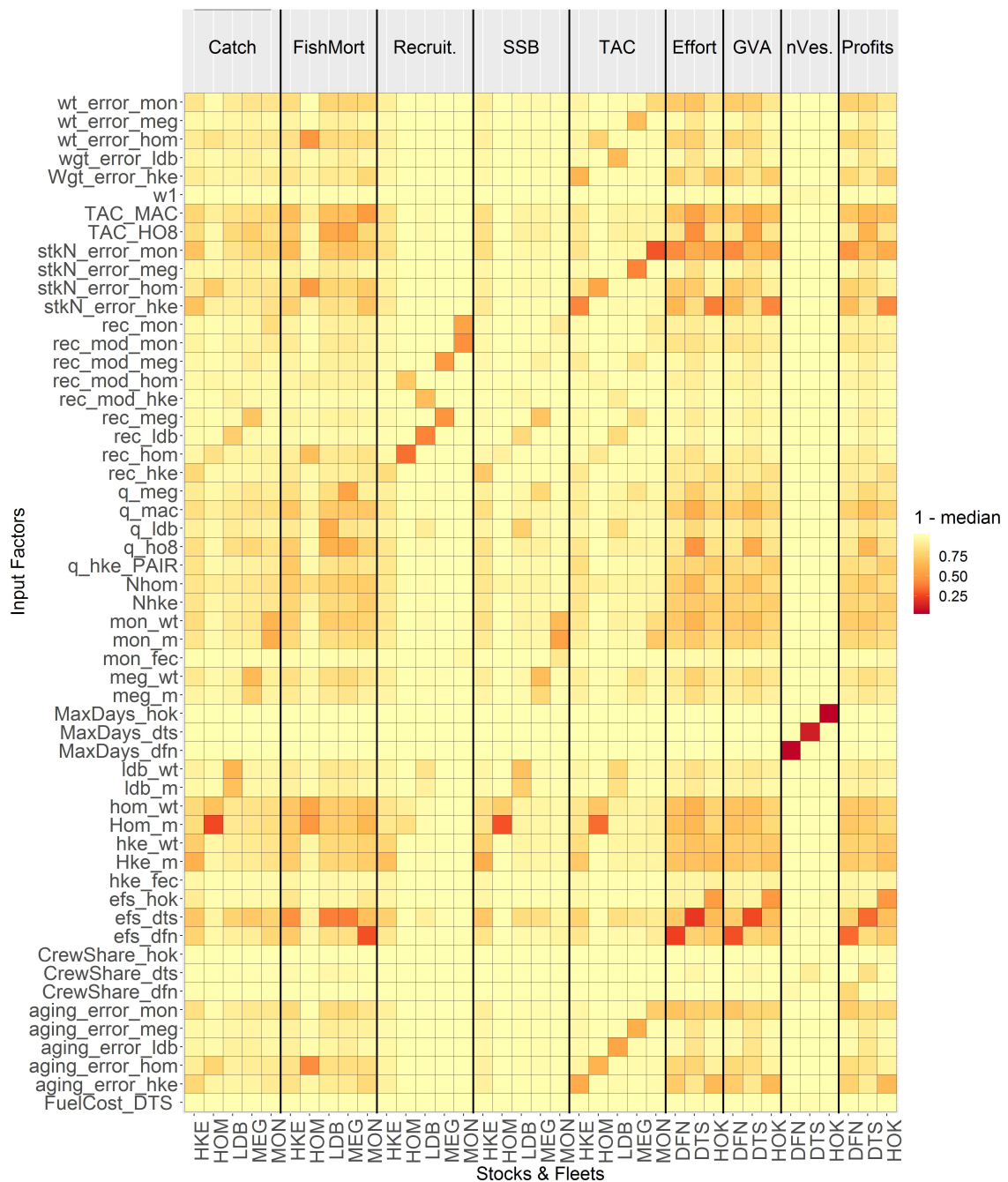


Figure 5: Variance explained by *total-effects* for all the output variables in 2020 year. Light yellow represents 0 and dark red 1.

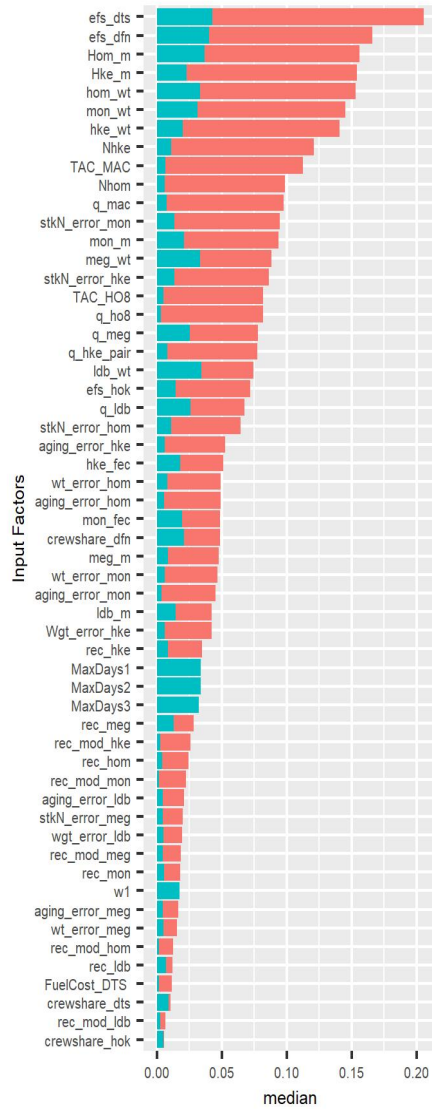


Figure 6: Generalized global sensitivity indices obtained using the 37 output variables. Blue bars correspond with the value of the *first-order* indices and the red bars with the difference between the *total-effects* and *first-order* indices. Therefore, the full bar shows the *total-effects*.

with this choice we make the output variables of which the variance is explained by few input factors predominant.

The performance indicator of the calibrated visual criterion is always closer to one than that of fixed-number of factors criterion (Table 4), i.e, the input factors selected with the calibrated visual criterion correspond with input factors that are higher in the ranking of the *total-effects*. The indicator for Savage criterion is the indicator closest to one only for the indicator at overall level when $Z \neq 4$ (Table 4).

Table 4: The performance indicator that measures the match between the ranking obtained in the generalized sensitivity indices and the indices selected by the Morris method using the calibrated visual criterion, the fixed number criterion and the Savage scores. The first column corresponds with the number of output variables used, the second column with the number of input factors, and the rest of the columns with the value of the performance indicator defined in equation 18 and the generalized performance index for each of the criteria.

| Output Variables | Input Factors | Performance indicator | | | Generalized Performance Indicator | | |
|------------------|---------------|-----------------------|--------|------------|-----------------------------------|--------|------------|
| | | Fixed Number | Savage | Calibrated | Fixed Number | Savage | Calibrated |
| 21 | 29 (Z=2) | 0.56 | 0.52 | 0.64 | 0.47 | 0.68 | 0.56 |
| 29 | 31 (Z=2) | 0.59 | 0.60 | 0.69 | 0.55 | 0.71 | 0.64 |
| 37 | 32 (Z=2) | 0.59 | 0.62 | 0.66 | 0.56 | 0.72 | 0.63 |
| 21 | 39 (Z=3) | 0.69 | 0.63 | 0.78 | 0.61 | 0.79 | 0.73 |
| 29 | 41 (Z=3) | 0.70 | 0.72 | 0.81 | 0.67 | 0.82 | 0.77 |
| 37 | 44 (Z=3) | 0.73 | 0.76 | 0.84 | 0.72 | 0.84 | 0.81 |
| 21 | 46 (Z=4) | 0.82 | 0.71 | 0.91 | 0.72 | 0.84 | 0.85 |
| 29 | 50 (Z=4) | 0.87 | 0.79 | 0.94 | 0.83 | 0.86 | 0.92 |
| 37 | 53 (Z=4) | 0.86 | 0.81 | 0.96 | 0.83 | 0.87 | 0.94 |

4. Discussion

We have defined a selection criterion for the Morris elementary effects method that allows to select the most important input factors using a cri-

terion that mimics the visual selection. Ideally, the selection should be done visually. However, the visual selection is not easily applied consistently when the number of output variables is high and the discrimination among input factors is unclear. Furthermore, it cannot be applied in an automatic way, for example in bootstrap simulations. The new criterion defined here provides a good approximation of the visual approach and has the advantages of being consistent along the whole selection process and of being able to be used in an automatic way. Other authors use the fixed-number of factors criterion applied to each output variable (Hussein et al., 2011; Morris et al., 2014; DeJonge et al., 2012). This approach is consistent along output variables, but could lead to unimportant input factors being selected in some cases (for example, in recruitment) and to important ones being discarded in others (for example, in profits). Campolongo et al. (2007) use Savage scores (Savage, 1956) to identify the most important input factors in a multi-dimensional output model. However, Savage scores are mostly used to compare ranking of input factors obtained using different approaches (Confalonieri et al., 2010; Borgonovo et al., 2003; Cucurachi et al., 2016) and their performance as a selection criterion has never been evaluated.

The calibrated visual criterion is better than fixed-number of factors and Savage criteria when comparing their performance for each output variable. Hence, if the objective is to explain the variance of every single output variable the calibrated visual criterion would be always preferred. For example, in the case study used here, the Savage criterion discards the input factor that explains most of the variance in the number of vessels

output variable. This happens because Savage criterion penalizes the input factors that are important in only one output variable in favor of those that are important in several variables, even if the variables are correlated. However, at overall level, if a small number of input factors are selected, the performance of the Savage criterion is better. This is because the basis of the generalized sensitivity index is more similar to the Savage criterion than to the calibrated visual one.

In summary, if the interest is to select the input factors that are the most important at overall level, even if the input factors that explain a significant part of the variance of a single output variable are left out and the number of input factors to be selected is low, the Savage criterion would be preferable. However, if the focus is on explaining the variance of every single output variable or the number of input factors to be selected is high, the calibrated visual criterion would be better.

The performance of the criteria has been evaluated using the ranking of the *total-effects* estimated by the Sobol method, considered as the reference method by many authors (Yang, 2011; Confalonieri et al., 2010; Sarrazin et al., 2016; Homma and Saltelli, 1996), for the input factors selected by one of the criteria evaluated here, the calibrated visual criterion. This fact may seem to produce a positive bias towards this criterion. However, the number of input factors selected by the criteria in the evaluation are lower than those considered in the Sobol method, especially for $Z < 4$. Hence, the ranking used for the performance evaluation is considered sufficiently broad to provide an unbiased assessment.

We select Lamboni et al.'s (2011) method to calculate the multivari-

ate indices because of its simplicity and ease of application. We discard Garcia-Cabrejo and Valocchi's (2014) method, because it requires adjusting a metamodel based on the polynomial chaos expansion. The most recent method, Xu et al.'s (2018), uses an index to assess the inputs' effect on the entire joint probability distribution of the multivariate output but its application is complex.

In the convergence assessment of the Morris method we take 95% as a threshold to ensure convergence. Nevertheless, other values could also be adequate, considering that higher values slow the convergence and lower values could lead to the selection of unimportant input factors. We recommend to use high values of alpha, as long as computational resources allow it.

We could have assessed the convergence using the "factor screening" criterion in Sarrazin et al. (2016). This criterion focuses on the width of the confidence interval of the non-selected input factors (input factors X for which $m_X^r < 0.95 \cdot N_{boot}$) and considers that it has converged when the width is narrower than 0.05. Of the 77 input factors with $m_X^r < 0.95 \cdot N_{boot}$, only 26, i.e., 34%, have converged when $r = 300$. Therefore, according to this criterion we should increase r with the subsequent increase in computational cost.

5. Conclusions

We have defined a selection and a convergence criteria to ensure a robust combination of the Morris method with the Sobol method or other GSA methods with a high computational cost. The calibrated visual crite-

riterion mimics the visual selection criterion, combining three of the features that are considered when selecting the input factors visually: the value of the absolute elementary effects in relation to the maximum, the number of input factors selected, and the difference in the absolute elementary effects between consecutive input factors. The criterion provides an objective method to select the most important input factors and a procedure to automatize the process. The automation allows its use in simulation mode, which is essential for calculating the confidence intervals of the indices using bootstrapping. In a comparison of the performance of three selection criteria the visual calibrated criterion has been the best of the three. The convergence criterion has been specifically defined to ensure the convergence of the Morris method when the objective is to select a maximum number of input factors. Moreover, the computing load required to achieve convergence for this criterion has proved to be lower than the criterion focused in the width of the confidence intervals.

Software availability

The R functions to implement the selection and convergence criteria are available in Zenodo <https://zenodo.org/record/3402534> (Garcia, 2019). In the same repository, there is an example showing how to use these functions in practice.

Acknowledgments

Financial support for this study was provided, in part, by the Department of Agriculture, Fishing and Food from the Basque Government

(FLBEIA and IMPACPES Grant agreement no. 227390 and 289257, respectively), the Department of Education, Language Policy and Culture from the Basque Government (IT1294-19) and BERC.2018-2021 program), the Spanish Ministry of Economy and Competitiveness MINECO and FEDER (MTM2016-74931-P, and BCAM Severo Ochoa excellence accreditation SEV-2017-0718). We would like to thank the two anonymous reviewers for their detailed comments and suggestions that have helped to greatly improve the quality of the document and Editage [<http://www.editage.com>] for editing and reviewing this manuscript for English language. This is publication number 936 of AZTI.

References

- Balci, O., 1997. Verification validation and accreditation of simulation models, in: Proceedings of the 29th Conference on Winter Simulation, IEEE Computer Society, Washington, DC, USA. pp. 135–141.
- Borgonovo, E., Apostolakis, G., Tarantola, S., Saltelli, A., 2003. Comparison of global sensitivity analysis techniques and importance measures in psa. *Reliability Engineering & System Safety* 79, 175 – 185. SAMO 2001: Methodological advances and innovative applications of sensitivity analysis.
- Borgonovo, E., Plischke, E., 2016. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research* 248, 869 – 887.
- Borgonovo, Emanuel Castaings, W.T.S., 2011. Moment independent im-

- portance measures: New results and analytical test cases. *Risk Analysis* 31, 404–428.
- Campbell, K., McKay, M.D., Williams, B.J., 2006. Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety* 91, 1468 – 1472. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004).
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software* 22, 1509–1518.
- Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. a unified approach. *Computer Physics Communications* 182, 978–988.
- Cariboni, J., Campolongo, F., 2004. Grouping model input factors to perform a sensitivity analysis computationally efficient, in: Spitzer, C., Schmocker, U., Dang, V.N. (Eds.), *Probabilistic Safety Assessment and Management*, Springer London, London. pp. 2018–2023.
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., Acutis, M., 2010. Comparison of sensitivity analysis techniques: A case study with the rice model warm. *Ecological Modelling* 221, 1897–1906.
- Cucurachi, S., Borgonovo, E., Heijungs, R., 2016. A protocol for the global sensitivity analysis of impact assessment models in life cycle assessment. *Risk Analysis* 36, 357–377.

- DeJonge, K.C., Ascough, J.C., Ahmadi, M., Andales, A.A., Arabi, M., 2012. Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling* 231, 113 – 125.
- Devroye, L., 1986. *Non-Uniform Random Variate Generation*. Springer-Verlag New York Inc.
- Garcia, D., 2019. dorleta/robust_Morris_Sobol: First release of Robust combination of Morris and Sobol methods.
- Garcia, D., Prellezo, R., Sampedro, P., Da-Rocha, J.M., Castro, J., Cerviño, S., García-Cutrín, J., Gutiérrez, M.J., 2017a. Bioeconomic multi-stock reference points as a tool for overcoming the drawbacks of the landing obligation. *ICES Journal of Marine Science* 74, 511–524. [10.1093/icesjms/fsw030](https://doi.org/10.1093/icesjms/fsw030).
- Garcia, D., Sánchez, S., Prellezo, R., Urtizberea, A., Andrés, M., 2017b. Flbeia : A simulation model to conduct bio-economic evaluation of fisheries management strategies. *SoftwareX* 6, 141–147.
- Garcia-Cabrejo, O., Valocchi, A., 2014. Global sensitivity analysis for multivariate output using polynomial chaos expansion. *Reliability Engineering & System Safety* 126, 25 – 36.
- Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety* 52, 1 – 17.

- Hussein, C., Verdoit-Jarraya, M., Pastor, J., Ibrahim, A., Saragoni, G., Pelletier, D., Mahévas, S., Lenfant, P., 2011. Assessing the impact of artisanal and recreational fishing and protection on a white seabream (*diplodus sargus sargus*) population in the north-western mediterranean sea, using a simulation model. part 2: Sensitivity analysis and management measures. *Fisheries Research* 108, 174–183.
- Lamboni, M., Monod, H., Makowski, D., 2011. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety* 96, 450–459.
- Morris, D.J., Speirs, D.C., Cameron, A.I., Heath, M.R., 2014. Global sensitivity analysis of an end-to-end marine ecosystem model of the north sea: Factors affecting the biomass of fish and benthos. *Ecological Modelling* 273, 251–263.
- Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33, 161–174.
- Norton, J., 2015. An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software* 69, 166 – 174.
- Norton, J.P., 2009. Selection of morris trajectories for initial sensitivity analysis. *IFAC Proceedings Volumes* 42, 670 – 674. 15th IFAC Symposium on System Identification.
- Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A sys-

- tematic review with practical workflow. *Environmental Modelling & Software* 79, 214 – 232.
- Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software* 67, 1 – 11.
- Rabitz, H., 1989. Systems analysis at the molecular scale. *Science* 246, 221–226.
- Reeves, S.A., 2003. A simulation study of the implications of age-reading errors for stock assessment and management advise. *ICES Journal of Marine Science* 60, 314–328.
- Ruano, M., Ribes, J., Seco, A., Ferrer, J., 2012. An improved sampling strategy based on trajectory design for application of the morris method to systems with many input factors. *Environmental Modelling & Software* 37, 103 – 109.
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling* 90, 229 – 244.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications* 181, 259–270.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity Analysis: The Primer*. Wiley.

- Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15, 377–395.
- Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global sensitivity analysis of environmental models: Convergence and validation. *Environmental Modelling & Software* 79, 135 – 152.
- Savage, I.R., 1956. Contributions to the theory of rank order statistics-the two-sample case. *Ann. Math. Statist.* 27, 590–615.
- Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. *Trends in Ecology & Evolution* 25, 479–486.
- Sobol, I., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* 7, 86 – 112.
- Sobol, I.M., 1993. Sensitivity analysis for nonlinear mathematical models. *Mathematical modeling and computational experiment* 1, 407–414.
- Xu, L., Lu, Z., Li, L., Shi, Y., Zhao, G., 2018. Sensitivity analysis of correlated outputs and its application to a dynamic model. *Environmental Modelling & Software* 105, 39–53.
- Yang, J., 2011. Convergence and uncertainty analyses in monte-carlo based sensitivity analysis. *Environmental Modelling & Software* 26, 444–457.

Appendix A. List of input factors

Table A.5: All the random factors considered in the GSA. They are ordered in alphabetical order. The name column correspond with the name used to denote the factors in the figures and the tables. The other three columns correspond with the component of the model the factor belongs to, its description and the stock and/or fleet it belongs to.

| Name | Component | Description | Stock/Fleet |
|---|-------------------------------------|--|---|
| AgingError_HKE AgingError_HOM AgingError_LDB AgingError_MEG AgingError_MON | Observation Model | Error in the aging process. The probability of assigning age 'i' to a fish of age 'j' | Hake H.Mackerel 4 Spot M. Megrin Monkfish |
| CapitalCost_DFN_SP CapitalCost_DTS_SP CapitalCost_HOK_SP | Fleets OM Entry-Exit Model | Current value of the capital invested multiplied with the opportunity cost of capital | Gillnetters Trawlers Longliners |
| Crewshare_DFN Crewshare_DTS Crewshare_HOK | Fleets OM Entry-Exit Model | The proportion of the turnover that is paid to the crew | Gillnetters Trawlers Longliners |
| DiscNError_HKE DiscNError_LDB DiscNError_MEG | Observation Model | Error in the observed numbers of discarded fishes | Hake 4 Spot M. Megrin |
| DiscWError_HKE DiscWError_LDB DiscWError_MEG | Observation Model | Error in the observed total weight of discards | Hake 4 Spot M. Megrin |
| Effshare_DFN Effshare_DTS Effshare_HOK | Fleets OM Short Term Dynamics | Distribution of total effort among metiers | Gillnetters Trawlers Longliners |
| Fcost_DFN Fcost_DTS Fcost_HOK | Fleets OM Entry-Exit Model | Fixed Cost per vessel Also used to calculate profits at fleet level | Gillnetters Trawlers Longliners |
| FuelCost_DFN_SP FuelCost_DTS_SP FuelCost_HOK_SP | Fleets OM Entry-Exit Model | Fuel Cost per unit of effort | Gillnetters Trawlers Longliners |
| InvestShare | Fleets OM Entry-Exit Model | Proportion of profits used to invest in new vessels | Fleet independent |
| LandNError_HKE LandNError_HOM LandNError_LDB LandNError_MEG LandNError_MON | Observation Model | Error in the observed numbers of landed fishes | Hake H.Mackerel 4 Spot M. Megrin Monkfish |
| LandWtError_HKE LandWtError_HOM LandWtError_LDB LandWtError_MEG LandWtError_MON | Observation Model | Error in the observed total weight of landings | Hake H.Mackerel 4 Spot M. Megrin Monkfish |
| M_HKE M_HOM M_LDB | Biological OM | Instantaneous rate of natural mortality at age | Hake H.Mackerel 4 Spot M. |

| | | | |
|---|--|--|---|
| M.MEG M.MON | | | Megrim Monkfish |
| Mat.HKE Mat.HOM Mat.LDB Mat.MEG Mat.MON | Biological OM | Proportion of mature individuals at age | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| MaxDays.DFN.SP MaxDays.HOK.SP MaxDays.DTS.SP | Fleets OM Entry-Exit Model | Maximum Number of Days a vessel can operate within a year | Gillnetters Trawlers Longliners |
| N.HO8 N.MAC N.WHB | Biological OM | Number of fish at age along the simulation | Western H. Mac. Mackerel Blue Whiting |
| N.HKE N.HOM N.LDB N.MEG N.MON | Biological OM | Number of fish at age in the first year of the simulation | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| price.HKE price.HOM.HO8 price.MAC price.MEG.LDB price.MON price.WHB | Fleets OM Entry-Exit Model | Fleet and Metier Independent Price of fish per ton | Hake Horse Mackerels Mackerel Megrims Monkfish Blue Whiting |
| price.OTH.DTS.SP.M1 price.OTH.DTS.SP.M2 price.OTH.DTS.SP.M3 price.OTH.HOK.DFN.M1 price.OTH.HOK.DFN.M2 price.OTH.HOK.DFN.M3 price.OTH.HOK.DFN.M4 price.OTH.HOK.DFN.M5 price.OTH.HOK.DFN.M6 | Fleets OM Entry-Exit Model | Price per ton of the OTH stock. The composition of OTH depends on the metier and hence the price is metier dependent. | OTH Trawlers OTB.DEM OTH Trawlers OTB.PEL OTH Trawlers PTB OTH G&L Trammel net OTH G&L Hand Line OTH G&L Longline OTH G&L Gillnet ≥ 100 OTH G&L Gillnet 60-79 OTH G&L Gillnet 80-99 |
| q.HKE.Baka.SP q.HKE.DTS.PT q.HKE.GNs_60 q.HKE.GNs_80 q.HKE.LLS q.HKE.Pair q.HKE.PGP.PT q.HO8 q.HOM.PS.PT q.HOM.DFN.HOK q.HOM.DTS.PT q.HOM.DTS.SP q.HOM.PGP.PT q.HOM.PS.SP q.LDB q.MAC q.MEG q.MON.DTS.PT q.MON.DTS.SP q.MON.HOK.DFN q.MON.PGP.PT q.OTH.DTS.SP.M1 q.OTH.DTS.SP.M2 q.OTH.DTS.SP.M3 q.OTH.HOK.DFN.M1 q.OTH.HOK.DFN.M2 q.OTH.HOK.DFN.M3 q.OTH.HOK.DFN.M4 q.OTH.HOK.DFN.M5 q.OTH.HOK.DFN.M6 | Fleets OM Catch Production Model | Catchability per fish stock (Defined at metier/fleet level depending on data availability) | Hake SP Trawl Hake PT Trawl Hake G&L 60-79 Hake G&L 80-99 Hake G&L Longline Hake SP Trawl PTB Hake PT PGP W. Horse Mackerel H. Mackerel PT PS H. Mackerel G&L H. Mackerel PT Trawl H. Mackerel SP Trawl H. Mackerel PT PGP H. Mackerel SP PS 4 Spot Megrim Mackerel Megrim Monkfish PT Trawl Monkfish SP Trawl Monkfish Sp. G&L Monkfish PT PGP OTH Trawlers OTB.DEM OTH Trawlers OTB.PEL OTH Trawlers PTB OTH G&L Trammel net OTH G&L Hand Line OTH G&L Longline OTH G&L Gillnet ≥ 100 OTH G&L Gillnet 60-79 OTH G&L Gillnet 80-99 |

| | | | |
|--|--|--|---|
| q_WHB | | | Blue Whiting |
| ret_HKE_DTS_SP ret_HKE_DTS_PT ret_LDB_DTS_SP ret_MEG_DTS_SP ret_MAC | Fleets OM Entry-Exit Model | The retention ogive. A vector at age with the proportion of catch that is retained onboard | Hake SP Trawl OTB Hake PT Trawl 4 Spot Megrim SP Trawl 5 Spot Megrim SP Trawl Mackerel all fleets |
| SR_params_HKE SR_params_HOM SR_params_LDB SR_params_MEG SR_params_MON | Biological OM Stock Recruitment Model | The parameters of the stock recruitment models | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| SR_uncerta_HKE SR_uncerta_HOM SR_uncerta_LDB SR_uncerta_MEG SR_uncerta_MON | Biological OM Stock Recruitment Model | A time series with the annual deviations of recruitment from stock-recruitment model | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| StkNError_HKE StkNError_HOM StkNError_LDB StkNError_MEG StkNError_MON | Observation Model | A vector at age with the observation error in the stocks numbers | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| StkWError_HKE StkWError_HOM StkWError_LDB StkWError_MEG StkWError_MON | Observation Model | A vector at age with the observation error in the stocks weight | Hake H.Mackerel 4 Spot M. Megrim Monkfish |
| TAC_HO8 TAC_MAC TAC_WHB | Advice Model | The TAC of the widely distributed stocks | Western H. Mac. Mackerel Blue Whiting |
| vcost_DFN vcost_DTS vcost_HOK | Fleets OM Entry-Exit Model | Variable Cost per unit of effort | Western H. Mac. Mackerel Blue Whiting |
| w1 w2 | Fleets OM Entry-Exit Model | Proportion in which capacity (in/de)crease (w1/w2) yearly | Fleet independent |
| Wt_HKE Wt_HOM Wt_LDB Wt_MEG Wt_MON | Biological OM Stock Recruitment Model | A vector at age with the mean weight of the fish individuals | Hake H.Mackerel 4 Spot M. Megrim Monkfish |

Appendix B. Notation

Notation

A, B The *sample* and *re-sample* matrices used to compute the importance indices in Sobol method.

$A_i, B_i, C_{k,i}$ The i -th row of the corresponding matrix.

C_k The matrix that is equal to A except in the column(s) that correspond with the k -th input factor, that is (are) taken from B matrix.

\mathbb{F} The set of all the input factors.

\mathbb{F}_r The set of input factors selected with morris method when r trajectories are used.

\mathbb{F}_D The set of input factors selected with the *factors distinguished from the others* criterion.

\mathbb{F}_F The set of input factors selected using the *fixed-number of factors* criterion.

\mathbb{F}_H The set of input factors selected with the *factors with high AEE value* criterion.

\mathbb{F}_V The set of input factors selected with the visual procedure.

\mathbb{F}_W The set of input factors selected with the weighted criterion.

\mathbb{F}_M The set of input factors selected when the Morris method is applied.

\mathbb{G}_T The set of total generalized indices.

| | |
|-------------------|--|
| G_{T_k} | Total-effect generalized index of the k -th input factor. |
| J | The dimension of the output of the simulation model. |
| j | subscript used in output variables along the manuscript. |
| K | The number of input factors. |
| k | subscript used for input factors along the manuscript. |
| K_r | The cardinality of \mathbb{F}_r . |
| $K_{EE,Z}$ | The number of input factors selected with the fixed number criterion when Z input factors are selected for each output variable. |
| K_{EE} | The number of input factors chosen <i>a priori</i> to be selected with the Morris method to be considered in the Sobol method. |
| M | Number of input factors, without grouping, in the Sobol method. |
| m_X^r | The number of iterations in which a input factor X is selected in the bootstrap of the Morris method with r trajectories. |
| N | The base sample size in the Sobol method. |
| N_{boot} | Number of bootstrap iterations. |
| \mathbb{P} | A large enough set of trajectories defined in ω . |
| \mathbb{P}_r | The r trajectories within \mathbb{P} that provide the best coverage of ω . |
| p | A trajectory in ω that belongs to \mathbb{P} . |
| R | The cardinality of \mathbb{P} . |

| | |
|-------------------|---|
| r | The number of trajectories used in the Morris method. |
| r_{max} | Maximum number of trajectories used in the Morris method. |
| \mathcal{S}_T^j | The set of <i>total-effects</i> of output variable Y_j . |
| S_k | <i>First-order</i> index for the k -th input factor. |
| S_{T_k} | <i>Total-effect</i> index for the k -th input factor. |
| $S_{T_k}^j$ | <i>Total-effect</i> index for the k -th input factor and output variable Y_j . |
| w_D | The weight given to the <i>factors distinguished from the others</i> criterion in the computation of the calibrated visual criterion. |
| w_F | The weight given to the <i>fixed-number of factors</i> criterion in the computation of the calibrated visual criterion. |
| w_H | The weight given to the <i>factors with high AEE value.</i> criterion in the computation of the calibrated visual criterion. |
| $X_{\sim k}$ | A sampling point in ω or Ω conditioned in all the input factors except the k -th one. |
| X | Input factor X or sampling point in ω or Ω . |
| X_k | k -th input factor. |
| \mathbf{Y} | A multidimensional output variable. |
| Y | An unidimensional output variable. |

- Z Number of input factors selected for each indicator in the application of fixed-number of factors in the evaluation of the performance of the selection indicators.
- α The threshold used for proportion to select the important input factors in the bootstrap of the Morris methods.
- Δ The width of the subintervals in the Morris method.
- δ_D The proportion used in the *factors distinguished from the others* criterion to select those input factors that are aside of the rest.
- δ_F The number of input factors selected in the *fixed-number of factors* selection criterion.
- δ_H The proportion used in the *factors with high AEE value* selection criterion.
- Γ A set of output variables.
- $\pi_X^r(i)$ Indicator variable of input factor X to be selected in iteration i of the Morris method with r trajectories.
- φ The simulation model.
- Ω The existence domain of the simulation model.
- ω The $[0, 1]^K$ unit hypercube.
- ρ_k^j Auxiliar variable used to calculate the performance indicators, Θ and Θ_G , in the evaluation of the selection criteria corresponding to the k -th input factor and the j -th output variable.

Θ The first performance indicator.

Θ_G The generalized performance indicator.

Appendix C. Supplementary Material: Morris absolute elementary effects graphs

For each output variable and stock or fleet combination only the 15 input factors with the highest AEE are plotted. The vertical lines correspond to the different criteria used to select the input factors. Red and Blue solid lines correspond to the selection criterion defined here and the visual selection respectively. The green, pink and light blue lines correspond with the high AEE, fixed number of input factors and input factors distinguished from the others criteria respectively.

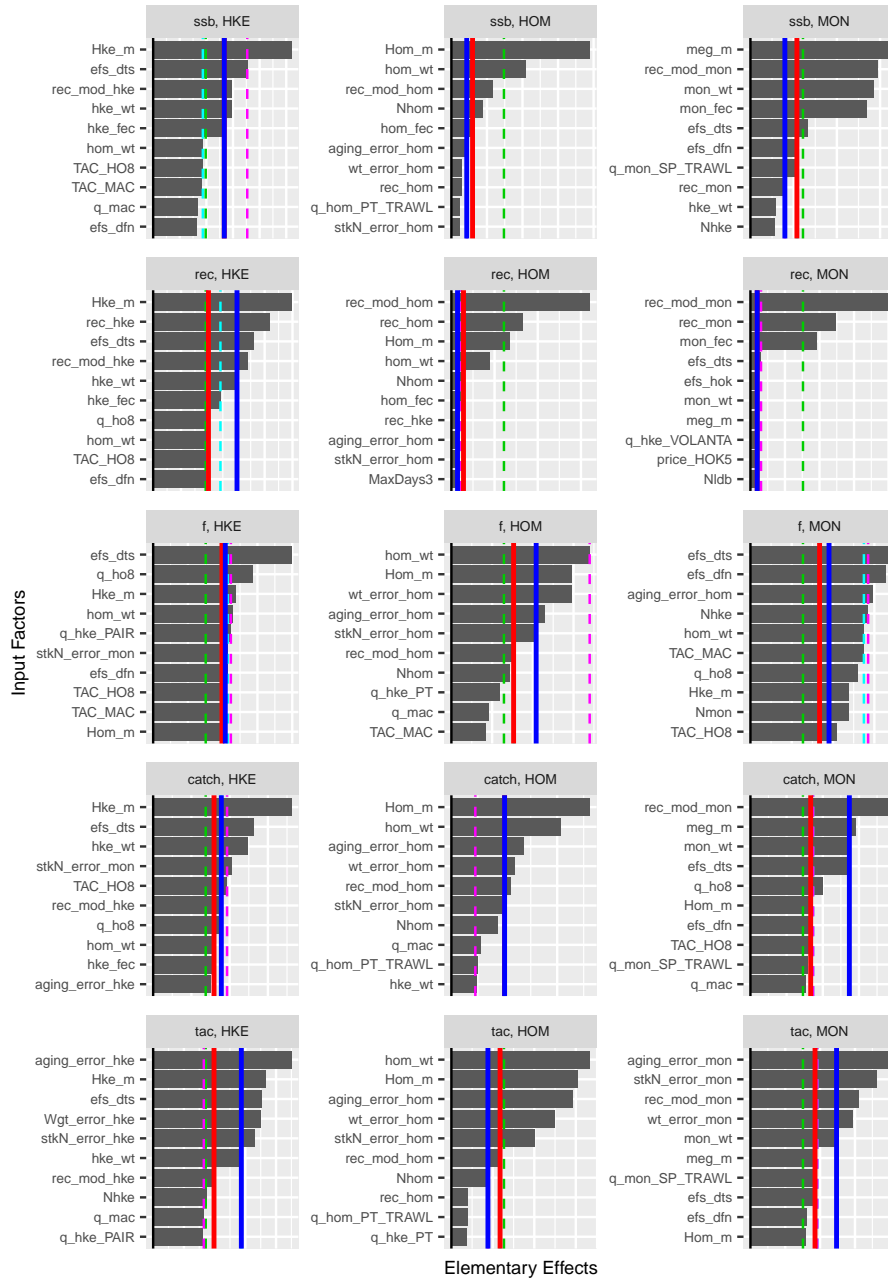


Figure S1: Absolute elementary effects (AEE) for hake (HKE), horse mackerel (HOM) and four spot megrim (LDB) stocks and all the stock output variables.

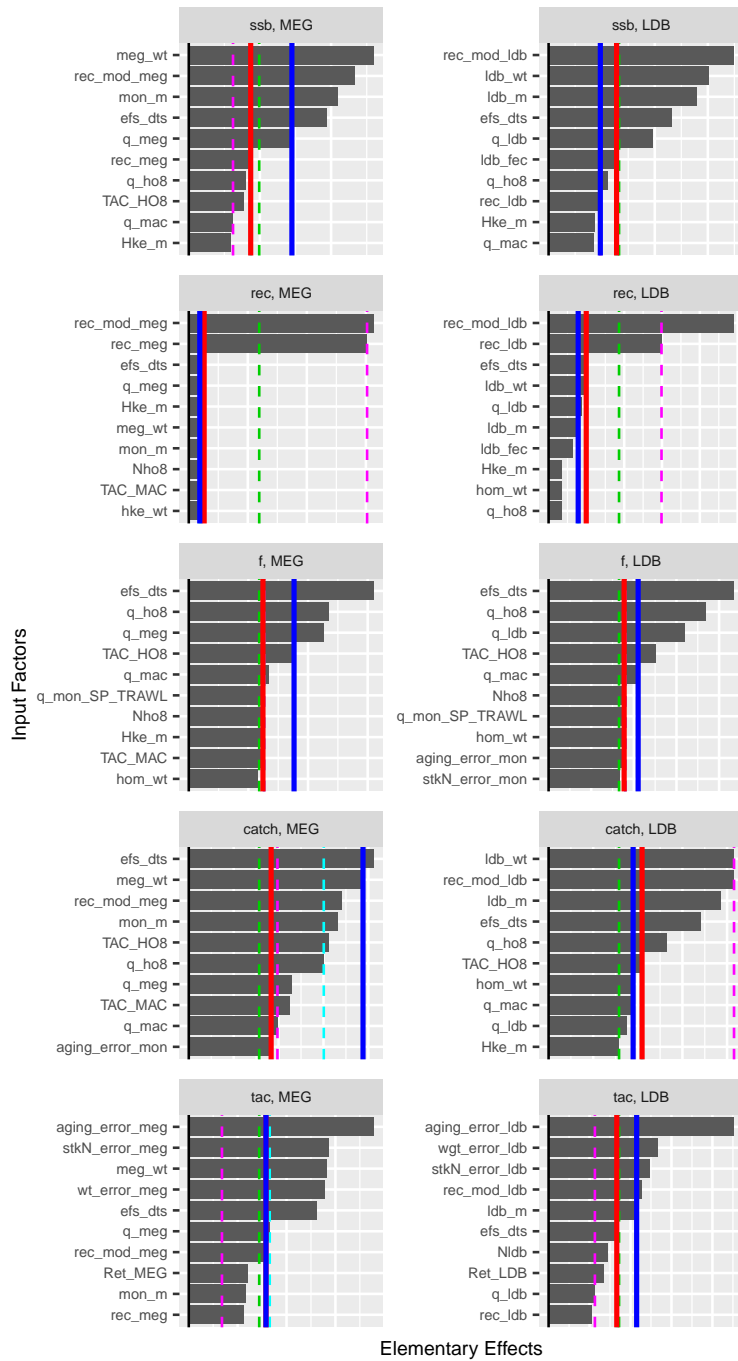


Figure S2: Absolute elementary effects (AEE) for megrim (MEG) and monkfish (MON) stocks and all the stock output variables.



Figure S3: Absolute elementary effects (AEE) for the spanish trawlers (DTS), gill-neters (DFN) and longliners (HOK) and all the fleet output variables.