



Test de puissance d'un échantillonnage

Comment savoir si un plan d'échantillonnage local permettra d'avoir des données exploitables localement avec le protocole STELI ?



Muséum
national
d'Histoire
naturelle



VIGIE NATURE

Avec le soutien de :



Programme coordonné par
la Société Française d'Odonatologie,
l'Office Pour les Insectes et leur Environnement,
le Conservatoire d'espaces naturels du Nord et du Pas-de-Calais et
le Muséum national d'Histoire naturelle

Dans le cadre du
Plan national d'actions en faveur des odonates
et du programme Vigie-Nature

L'idéal pour étudier la biodiversité serait de savoir avec certitude si telle espèce est présente ou absente d'un site et de connaître le nombre exact d'individus présents.

Cet idéal étant inaccessible (ou très difficilement), il faudra se contenter de probabilités et essayer de se rapprocher le plus possible des vraies valeurs à partir d'échantillons plus ou moins grands.

Les données récoltées avec le protocole STELI permettent de calculer des probabilités de présence (ψ) et de détection (p) des espèces de libellules contactées sur les sites suivis.

L'objectif du test de puissance développé ci-après est de connaître le nombre de sites et le nombre de passages minimum qu'il faudra effectuer pour obtenir la valeur la plus précise et la moins biaisée possible de ψ et de p .

Cette incertitude sera quantifiée par une variance qui représente la largeur de la distribution attendue autour de la valeur « vraie » (voir diapo ci-après)

Qu'est-ce qu'une variance ?

Comme nous ne pourrions jamais suivre chaque libellule présente en France pour savoir quelle espèce se porte bien ou mal, nous procédons à un suivi d'un échantillon du territoire français.

Les résultats obtenus via un échantillonnage étant une partie de la réalité, il est important de savoir à quel point on s'écarte de la réalité (sachant que plus un échantillonnage est important, plus on se rapproche de la réalité).

Une variance est une dispersion des valeurs autour d'une moyenne. Plus la variance est grande, plus l'incertitude est grande. Ainsi, la variance quantifie l'incertitude autour de la moyenne.

Avant toute chose, télécharger le logiciel SODA

“Single-season Occupancy study Design Assistant”

<http://www.kent.ac.uk/smsas/personal/msr/soda.html>

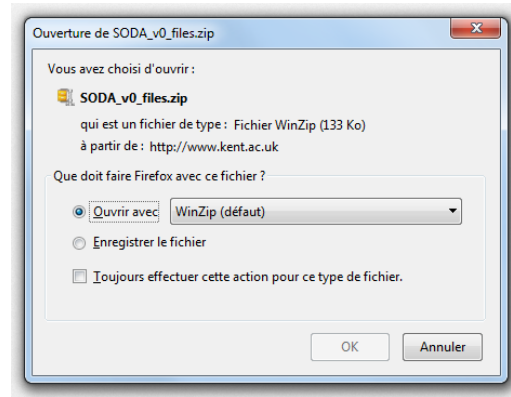
1 - Télécharger Matlab Compiler Runtime (MCR)

<http://www.kent.ac.uk/smsas/personal/msr/soda/MCRInstaller.exe>



2 - Puis télécharger le package SODA

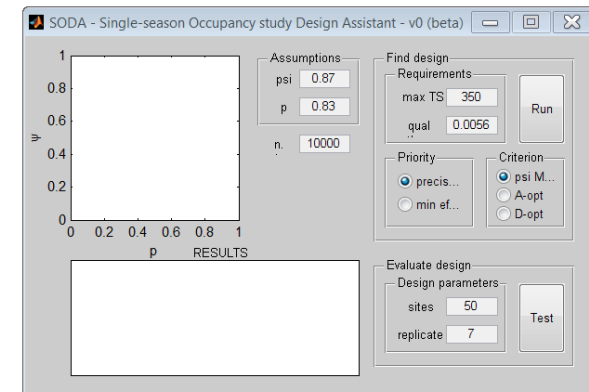
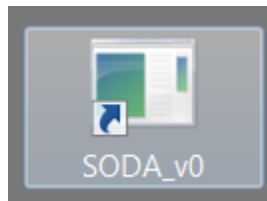
http://www.kent.ac.uk/smsas/personal/msr/webfiles/soda/SODA_v0_files.zip



Extraire le dossier et double cliquez sur le fichier SODA_v0.exe

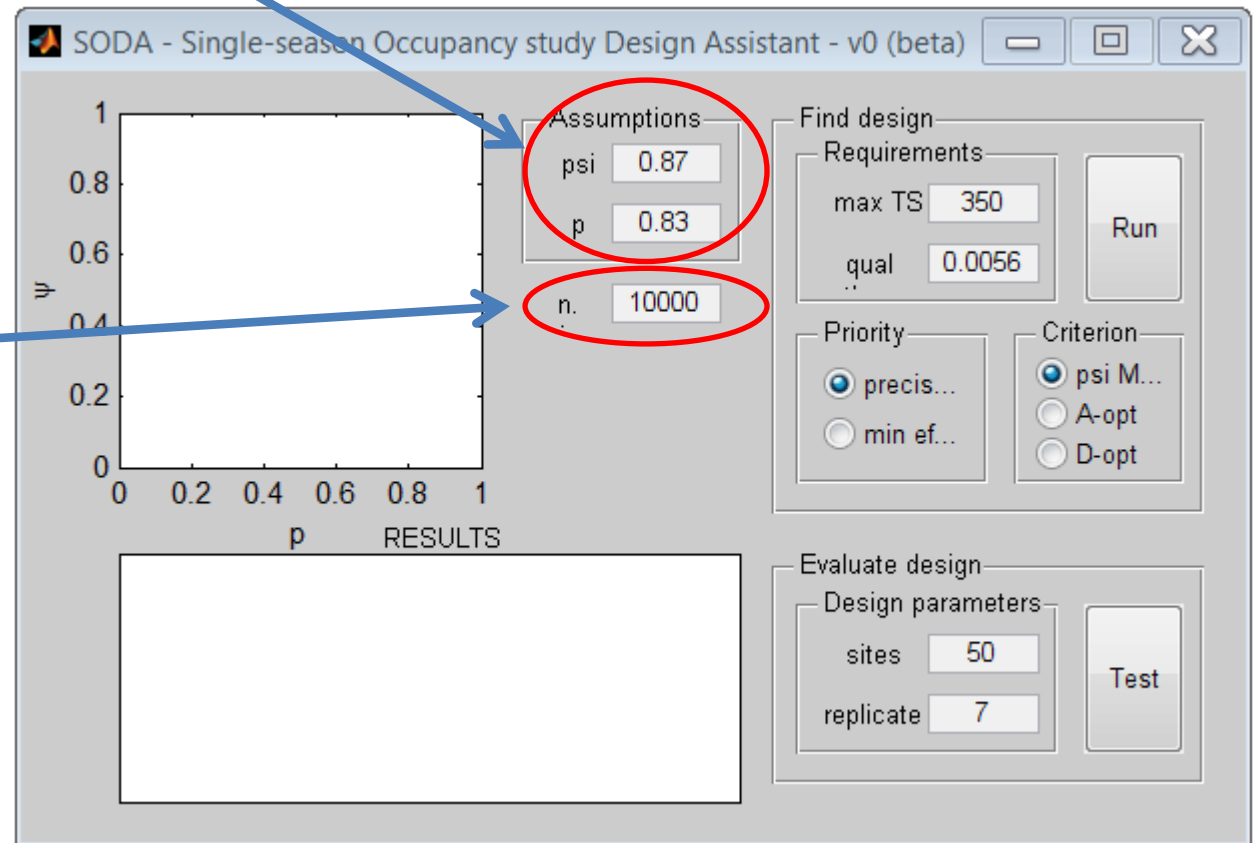
Et le logiciel s'ouvre !

Vous pouvez créer un raccourci sur votre bureau pour y accéder directement la prochaine fois.



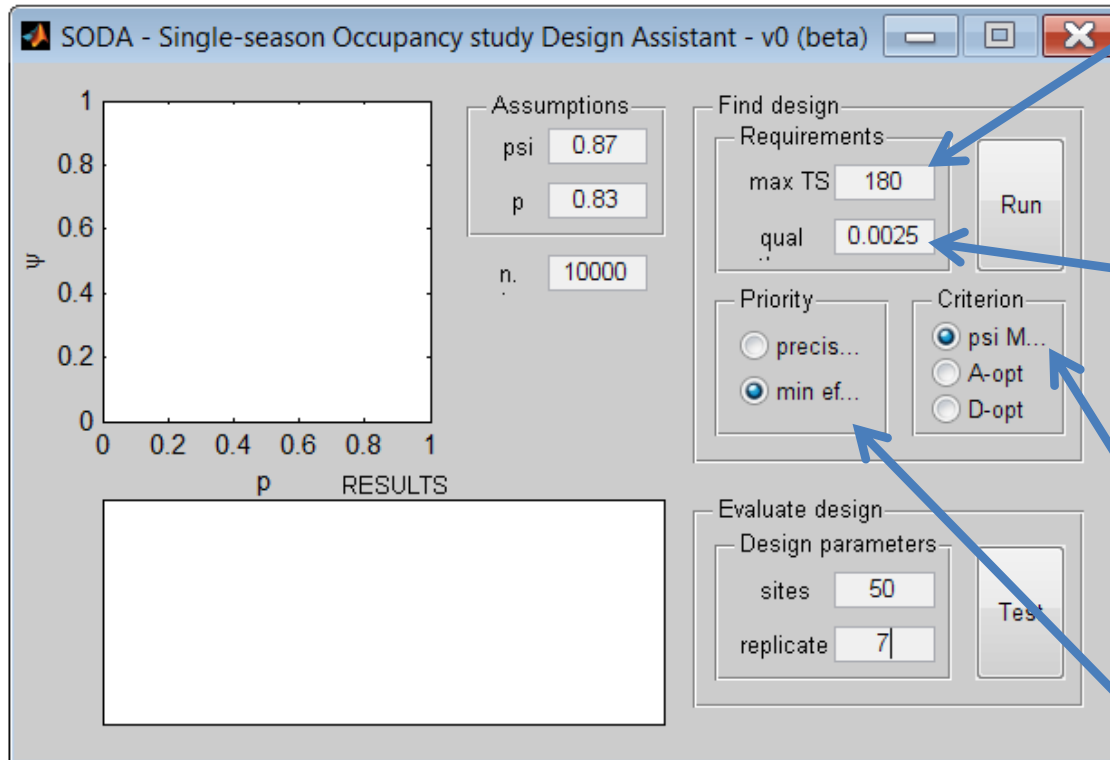
Pour commencer, entrer le taux d'occupation de l'espèce (ψ) et sa probabilité de détection (p).
Ces valeurs s'obtiennent suite à une première évaluation de ces paramètres calculés avec le logiciel Presence (voir le tutoriel d'utilisation de Presence). Les valeurs nationales obtenues grâce au STELI peuvent être utilisées (elles sont disponibles dans le bilan annuel).

Indiquer ici le nombre de simulations souhaitées (le nombre 10000 par défaut peut être gardé)



Exemple de *Ischnura elegans* : $\psi = 0.8658$ $p = 0.828$

La section « Find design » permet de d'évaluer le nombre de sites et de passages minimum qu'il faut prévoir, soit pour maximiser le calcul des paramètres, soit pour minimiser l'effort de terrain.



Max TS : c'est la quantité de terrain maximum réalisable. Il est évalué en multipliant le nombre de sites par le nombre passages. Par exemple, si on ne peut pas faire plus de 60 sites et 3 passages par site, on indiquera max TS = 180 (c'est-à-dire 60 x 3)

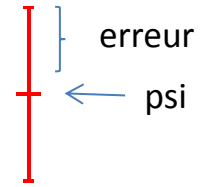
Qual, correspond à l'erreur qu'on accepte pour estimer psi.
Son attribution est expliquée dans la diapo suivante.

« Criterion » permet de sélectionner le type de valeur qu'on souhaite optimiser. Dans notre cas, on souhaite minimiser la variance du taux d'occupation (psi), on gardera alors « Psi MSE ».

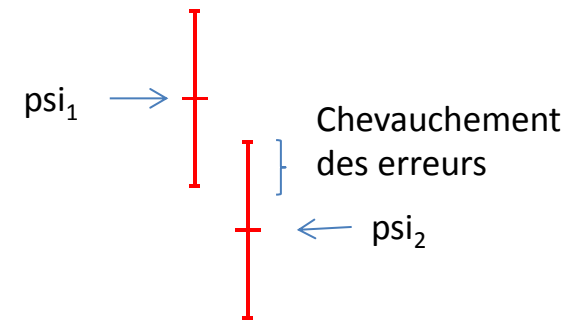
Priority, permet de définir si on préfère optimiser la précision des paramètres ou si on souhaite minimiser l'effort de terrain.

Comment définir l'erreur acceptable de psi ?

Le taux d'occupation des espèces (psi) est une estimation qui comporte une incertitude (erreur standard) qu'on peut schématiser ainsi :



Si l'amplitude de cette erreur est importante, elle peut empêcher la détection des variations de psi d'année en année. En d'autre terme, si l'erreur de $\psi_{(\text{année } 1)}$ chevauchent l'erreur de $\psi_{(\text{année } 2)}$, la variation annuelle aura peu de chances d'être détectée par les tests statistiques.



Ainsi, il faut définir pour chaque variation de psi supposée l'erreur maximale acceptable pour ne pas que les erreurs se chevauchent.

On partira du principe que l'erreur de $\psi_{(\text{année } 1)}$ est la même que l'erreur de $\psi_{(\text{année } 2)}$. Pour que ψ_1 soit vraiment différent de ψ_2 , il faut donc arriver à la logique suivante : $\psi_1 - \text{erreur} > \psi_2 + \text{erreur}$

Etant donnée que $\psi_2 = \psi_1 + x$ (x étant la variation de psi entre l'année 1 et l'année 2)

On peut écrire : $\psi_1 - \text{erreur} > \psi_1 + x + \text{erreur}$

Soit $\text{erreur} < x/2$

Pour SODA, Qual doit être le plus petit possible. On peut ainsi considérer que Qual doit être inférieur à $x/2$

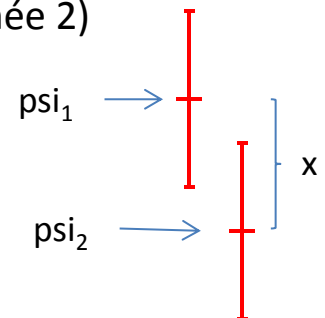


Tableau des correspondances sur l'écart attendu entre l'estimation de $\psi_{(année\ 1)}$ et $\psi_{(année\ 2)}$, et l'erreur acceptable (Qual)

Différence qu'on souhaite détecter entre ψ_{1} et ψ_{2} (x)	Erreur maximale acceptable (x/2)
0.01	0.005
0.02	0.01
0.03	0.015
0.04	0.02
0.05	0.025
0.06	0.03
0.07	0.035
0.08	0.04
0.09	0.045
0.10	0.05
0.11	0.055
0.12	0.06
0.13	0.065

Si ψ varie de 0,01 entre l'année 1 et l'année 2

alors, l'erreur acceptable de ψ (Qual) doit être au maximum de 0,005

alors on pourra détecter la variation de ψ si elle est supérieure à 0,10

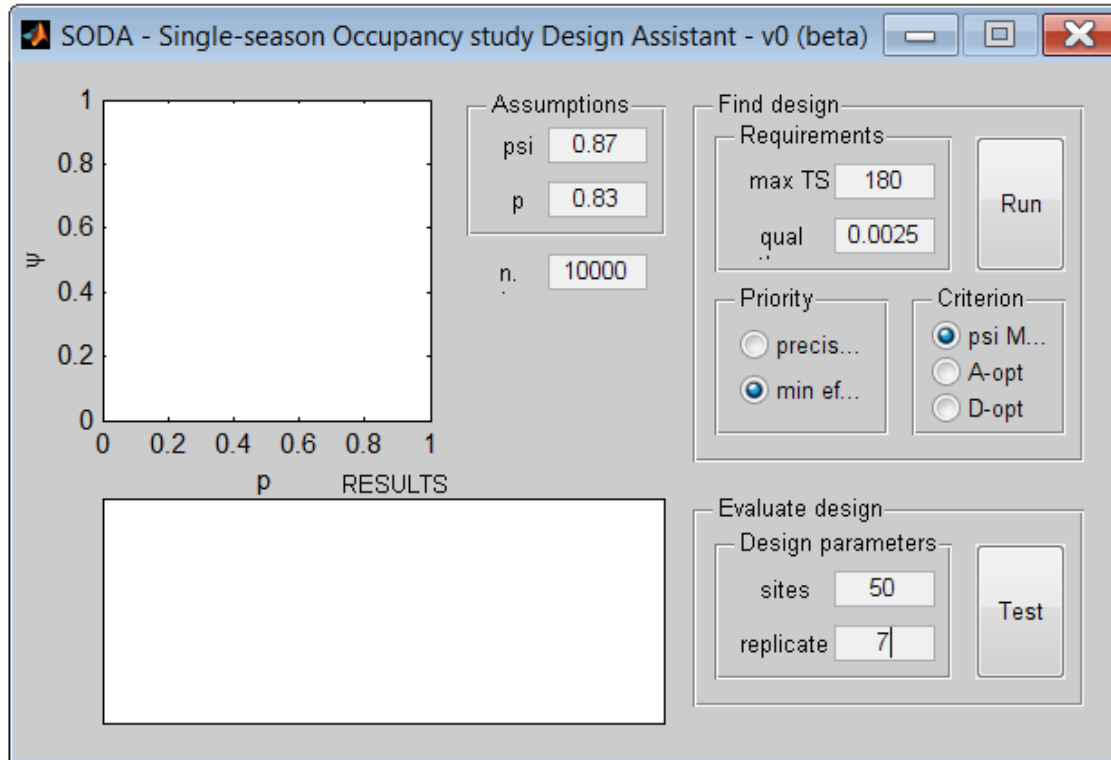
Si ψ à une erreur de 0,05

...

0.90	0.45
0.91	0.455
0.92	0.46
0.93	0.465
0.94	0.47
0.95	0.475
0.96	0.48
0.97	0.485
0.98	0.49
0.99	0.495
1.00	0.5

...

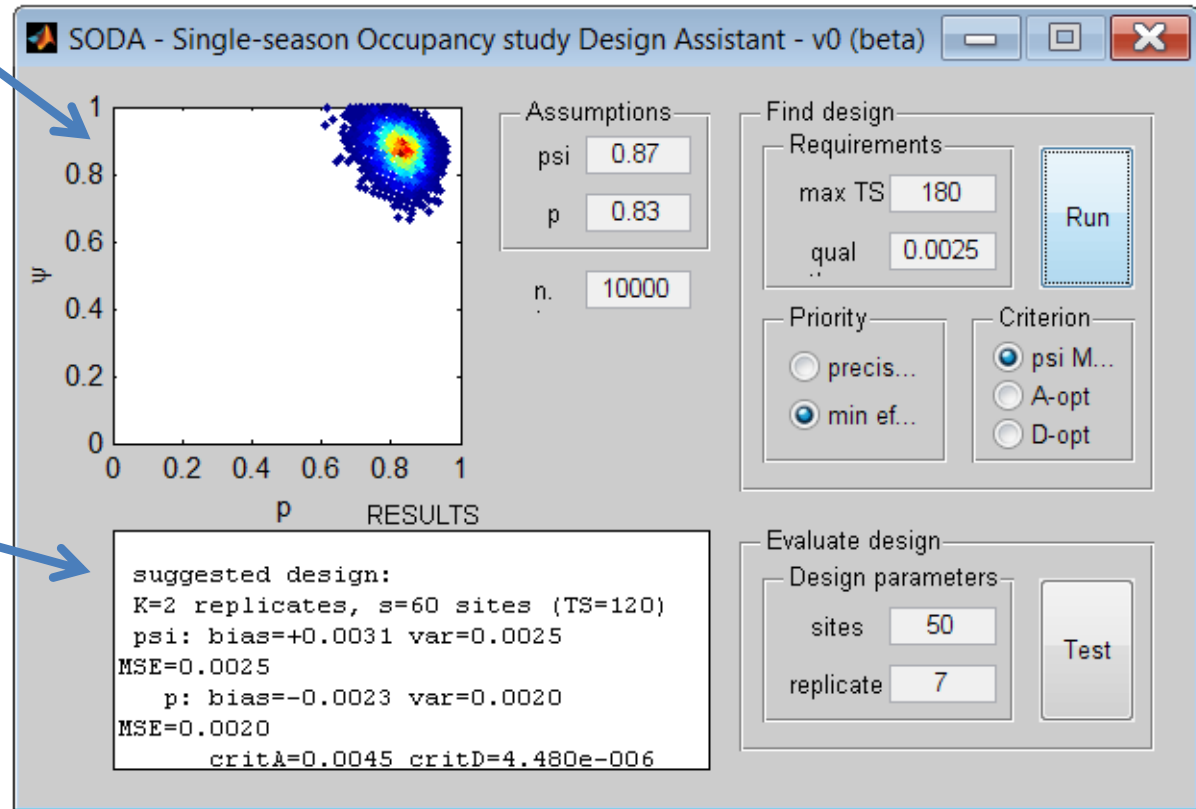
Une fois toutes les configurations effectuées, cliquer sur « Run » pour lancer les calculs



Le nuage de point donne une image de la précision des données qui seront récoltées selon les paramètres rentrés précédemment.

S'il est compact, la précision est bonne. S'il est très lâche, les biais seront importants. (voir diapo d'après pour voir des exemples)

Pour minimiser l'effort sur le terrain et garder une précision minimum de l'estimation du taux d'occupation (ψ), il est suggéré, pour *Ischnura elegans*, de réaliser un échantillonnage de 60 sites et de ne réaliser finalement que 2 passages par site.



Un autre exemple avec max TS = 20

The screenshot shows the SODA software interface. On the left, a scatter plot displays the relationship between ψ (y-axis, 0 to 1) and p (x-axis, 0 to 1). The plot is titled 'RESULTS' and shows a dense cluster of points, with a blue arrow pointing from the text box on the left to the plot area.

Assumptions:

- psi: 0.87
- p: 0.83
- n: 10000

Find design Requirements:

- max TS: 20 (circled in red)
- qual: 0.0025

Priority:

- precis...
- min ef...

Criterion:

- psi M...
- A-opt
- D-opt

Evaluate design Design parameters:

- sites: 50
- replicate: 7

RESULTS:

```
no design found to achieve targets!  
suggested design:  
K=2 replicates, s=10 sites (TS=20)  
psi: bias=+0.0051 var=0.0128  
MSE=0.0128  
p: bias=-0.0031 var=0.0114  
MSE=0.0114
```

Avec max TS = 20

Echantillonnage insuffisant pour la précision requise. Il faut soit augmenter l'échantillonnage soit baisser son niveau d'exigence pour l'estimation des paramètres.

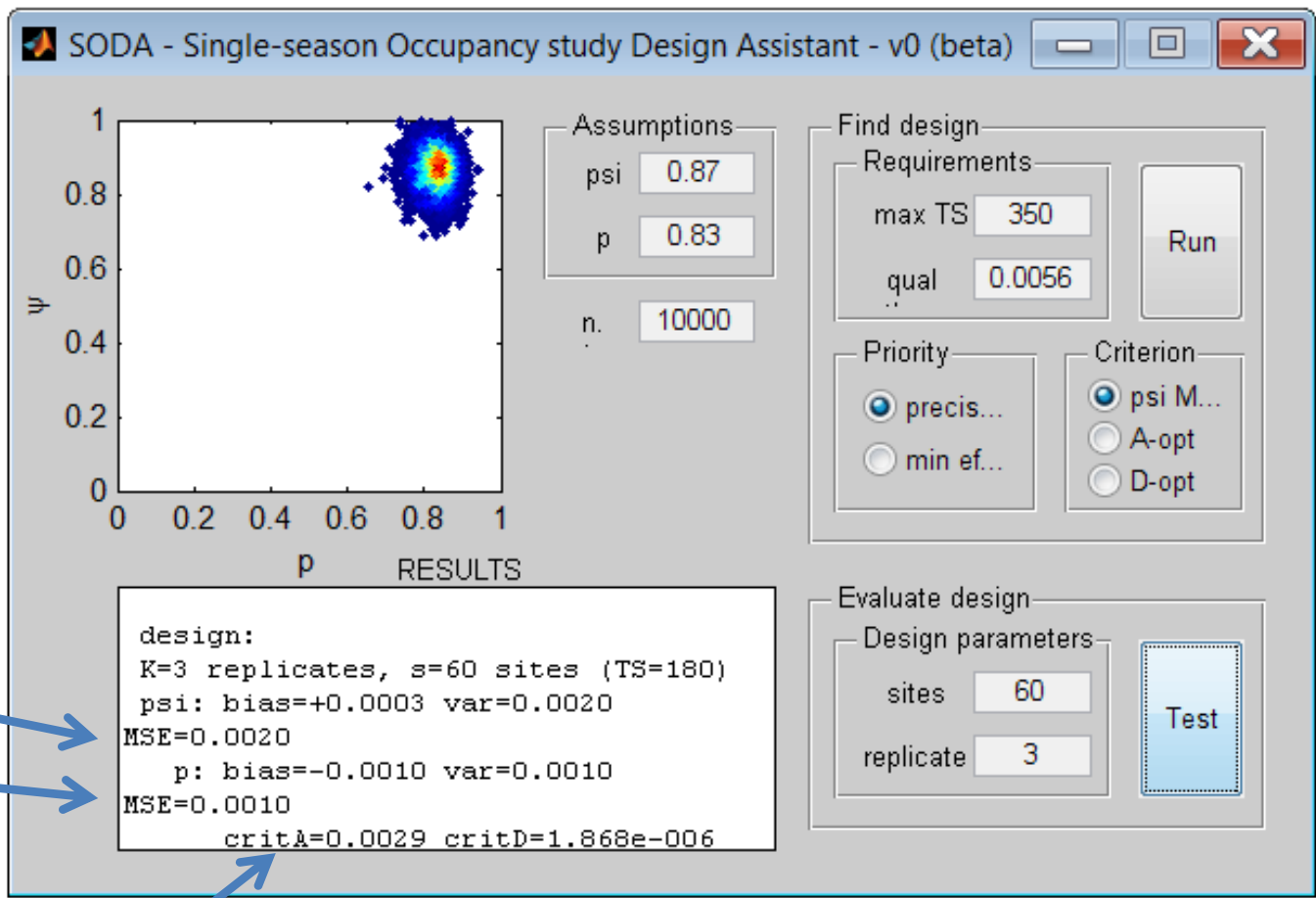
La section « Evaluate design » permet d'évaluer si un plan d'échantillonnage est suffisant pour estimer correctement Psi

Entrer le taux d'occupation de l'espèce (psi) et sa probabilité de détection (p).

The screenshot shows the SODA - Single-season Occupancy study Design Assistant - v0 (beta) window. It features a plot area on the left with axes labeled ψ (y-axis, 0 to 1) and p (x-axis, 0 to 1). Below the plot is a large empty box labeled "RESULTS". To the right of the plot is the "Assumptions" section with input fields for "psi" (0.87), "p" (0.83), and "n." (10000). Further right is the "Find design" section with "Requirements" (max TS: 350, qual: 0.0056) and a "Run" button. Below that is the "Priority" section with radio buttons for "precis..." (selected) and "min ef...". To the right of "Priority" is the "Criterion" section with radio buttons for "psi M..." (selected), "A-opt", and "D-opt". At the bottom right is the "Evaluate design" section with "Design parameters" (sites: 60, replicate: 3) and a "Test" button. A blue arrow points from the text above to the "psi" input field. Another blue arrow points from the text in the box on the right to the "sites" input field.

Le nombre de sites effectués et le nombre de réplicas par site

Cliquer sur « Test » pour lancer les calculs



Variance de Psi

Variance de P

```

design:
K=3 replicates, s=60 sites (TS=180)
psi: bias=+0.0003 var=0.0020
MSE=0.0020
    p: bias=-0.0010 var=0.0010
MSE=0.0010
        critA=0.0029 critD=1.868e-006
  
```

Somme de la variance de psi et p