

# Package VSURF

## Fontctionnement et possibilités

Le package VSURF a été imaginé par Robin Genuer (2015) et propose une méthode de sélection de variables basée sur des forêts aléatoires (Genuer et al., 2010).

Pour illustrer ses possibilités nous prenons le jeu de données “Toys” disponible nativement sur R.

Le package permet de réaliser la classification ou la régression étape par étape...

```
library(VSURF)
data("toys")
step1 <- VSURF_thres(toys$x, toys$y, mtry = 100)
step2 <- VSURF_interp(toys$x, toys$y, vars = step1$vselect.thres)
step3 <- VSURF_pred(toys$x, toys$y,
                    err.interp = step2$err.interp,
                    vselect.interp = step2$vselect.interp)
```

... mais également en une seule fonction.

```
vsurfToys <- VSURF(toys$x, toys$y, mtry = 100)
```

Les résultats se présentent ainsi sous la forme :

```
summary(vsurfToys)
```

VSURF computation time: 4.5 secs

VSURF selected:

- 24 variables at thresholding step (in 3.5 secs)
- 5 variables at interpretation step (in 0.9 secs)
- 3 variables at prediction step (in 0.1 secs)

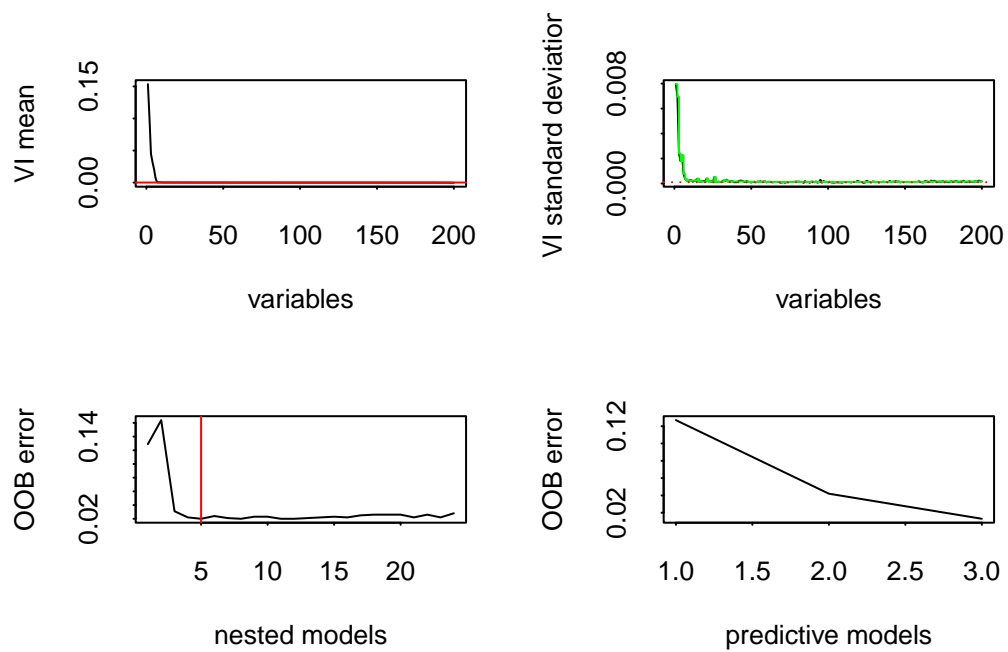
```
print(vsurfToys$varselect.interp)
```

```
[1] 3 2 6 5 1
```

```
print(vsurfToys$varselect.pred)
```

```
[1] 3 6 5
```

```
plot(vsurfToys)
```



```
plot(vsurfToys, step = "thres", imp.mean = FALSE, ylim = c(0, 2e-03))
```

