## PATApi : un nouvel outil de planification des pesées, rapide et performant

LEGRIS M. (1), DELPEUCH A. (1), AUGIER G. (2), JEANNOT L. (3), LE HUNG M. (3), DELGOULET D. (3) GRIFFON L. (3)
(1) Institut de l'Elevage, 149 rue de Bercy – 75 595 Paris, France
(2) Eliance, 149 rue de Bercy – 75012 Paris, France
(3) Races De France, 149 rue de Bercy – 75012 Paris, France

**RESUME**
L'élevage de bovins de races à viandes représente 3,7 millions de vaches en France en 2022 (1), dont 1 million avec pedigree certifié (CPB), et 435 000 en Contrôle de Performance Allaitant (CPV). Chaque année, plus d'un million de pesées de veaux sont réalisées jusqu'au sevrage, donnant 331 000 poids à âge type à 210 jours (PAT210). En 2021, le projet CALPAT a fourni de nouvelles règles de calcul des PAT associées à un indicateur de fiabilité (2). Ces évolutions ont permis d'augmenter le nombre de PAT mais ont accentué les difficultés déjà existantes pour les techniciens et les éleveurs à planifier les pesées. Pour résoudre ces problèmes, le projet PATApi, financé par France Génétique Elevage, a eu pour objectif de fournir un outil performant de planification des pesées jusqu'au sevrage. Ses principales fonctionnalités attendues par les conseillers sont : 1°) un outil de prévision, capable de fournir des dates de pesées théoriques basées sur les naissances de l'année précédente. 2°) des mises à jour régulières basées sur les naissances et les pesées réelles. 3°) une liste d'animaux avec dates de naissance et paramètres en entrée, 4°) en sortie, des dates de passage optimales et des éléments d'aide à la décision permettant des représentations graphiques et un ajustement dynamique par l'utilisateur. La recherche exhaustive de dates s'est imposée parmi les méthodes disponibles. Cependant, elle impliquait de considérer plus de 4 milliards de combinaisons possibles pour une distribution des naissances sur 4 mois. Une première phase d'optimisation a permis de détecter les périodes optimales plutôt que les dates exactes et de réduire considérablement le nombre de combinaisons testées. Finalement, la méthode stochastique du "recuit simulé", a fourni le nombre optimal de pesées et les dates de pesées correspondantes, avec une amélioration très significative des temps de calcul. En pratique et pour un élevage de 60 vaches avec 3 pesées, l'algorithme restitue les résultats en 3 à 5s. Une fois amélioré, cet algorithme, développé en langage R, a été packagé et encapsulé dans une API. Celle-ci interagira avec tout logiciel capable de transmettre une liste d'animaux et les paramètres d'entrée attendus. Son déploiement, prévu en 2024, doit faciliter le travail des techniciens et des éleveurs allaitants dès la campagne prochaine.

## New algorithm to plan easier and faster weighings for French breeders and technicians.

LEGRIS M. (1), DELPEUCH A. (1), AUGIER G. (2), JEANNOT L. (3), LE HUNG M. (3), DELGOULET D. (3) GRIFFON L. (3)
(1) Institut de l'Elevage, 149 rue de Bercy – 75 595 Paris, France

**SUMMARY**
Beef cattle breeding represent 3.7 million cows in France in 2022 (1), including 1 million with certified pedigree (CPB), and 435,000 in Beef Cattle Recording (BCR). Each year, more than 1 million calves were weighed up to weaning, giving 331,000 Adjusted Weights at 210 days (AW210). In 2021, the CALPAT project provided new flexible rules for calculating AW and a reliability indicator was developed (2). These recent developments increased the number of AWs but pointed the already existing difficulties for technicians and breeders to plan weighing sessions. To solve these problems, the PATApi project was adopted by « France Genetique Elevage », to provide a high-performance tool for planning weighings up to weaning.
At the beginning, we initially focused on the needs of network advisors with regard to weighing planning. The main features expected from the algorithm are: 1°) a forecasting tool, able to provide theoretical weight dates based on the previous year's births. 2°) regular updates based on actual births and weighings. 3°) a list of animals with dates of birth and parameters 4°) at the output, provide optimal passage dates, decision-support elements such as the ratio of animals with AWs, their average reliability, and graphical representations enabling dynamic adjustment by the user. At the same time, an exploratory analysis of available methods identified the exhaustive search for dates as a solution, particularly for providing dynamic graphical representations. However, this method would involve considering more than 4 billion possible combinations for a 4-month distribution of births and require 4 different visit dates to estimate AWs. An initial optimisation phase enabled us to detect the optimum periods (represented by their median date) rather than the exact dates, and to considerably reduce the number of combinations tested. Finally, the stochastic 'simulated annealing' method completely adapted to this problem, provided the optimum number of weighings and the corresponding weighing dates, with a very significant improvement in calculation times. On the downside, this method does not allow dynamic adjustment by the user. In practice, the performances obtained are satisfying between 3 and 5s for a farm of 60 animals with 3 weighing dates. Once improved, this algorithm, developed in R language, was packaged and encapsulated in an API. This will interact with any software able to transmit a list of animals and the expected input parameters. Commissioning is expected as from 2024 and it will facilitate the work of technicians and breeders alike from the start of the next birth campaign.

## INTRODUCTION

En 2023 en France, le Contrôle des Performances Viande (CPV) représentait plus de 6 000 éleveurs et près de 435 000 vaches en VA4 (Foucault M., 2023). Cette formule, qui associe pesée et pointage autour du sevrage, concerne environ 14 % des veaux nés en France. Environ 400 000 poids à âge type (PAT) en pré-sevrage sont calculés chaque année pour un peu plus d'un million de pesées, dont 80 % sont réalisées par des techniciens. Il faut en moyenne 2,5 pesées par animal pour obtenir au moins 1 PAT pré-sevrage (PAT 120 jours ou PAT 210 jours). Cela nécessite entre 2 et 4 visites par élevage selon l'étalement des naissances. Les PAT sont utilisées pour conseiller l'éleveur dans sa conduite et alimentent l'évaluation génétique (IBOVAL) pour le calcul des index de croissance (CRsev) et de valeur laitière (ALait).

Depuis une dizaine d'années, l'activité du CPV a connu une triple évolution :

- Les attentes des éleveurs en matière de conseil et service sont de plus en plus hétérogènes et différenciées selon les zones,
- Les exploitations se spécialisent, la taille des troupeaux augmente et la main d'œuvre se raréfie, ce qui tend à augmenter les contraintes de réalisation des pesées nécessaires à l'obtention des résultats techniques et génétiques,
- Les Organismes de Sélection (OS), maîtres d'œuvre du CPV, souhaitent que la collecte des données puisse répondre aux objectifs de sélection de leurs races, apportant une différenciation supplémentaire dans la gestion du processus.

## 1. MATERIEL ET METHODES

La planification des pesées dans un élevage donné vise à permettre l'obtention d'un maximum de PAT (120 jours, 210 jours ou les 2), les plus fiables possibles (Delpeuch A., 2021) avec le minimum de pesées par animal, selon la répartition des naissances. Un élevage ayant une période de naissance de 5 mois et souhaitant obtenir les 2 PAT avec un niveau de fiabilité permettant une évaluation génétique devra réaliser au moins 4 visites. A l'inverse et pour des PAT de même fiabilité, un élevage ayant des vêlages répartis sur 2 mois pourra s'en satisfaire de 2.

### 1.1. RECEUIL DES BESOINS ET DES ATTENTES DES UTILISATEURS

Afin d'identifier les attentes des utilisateurs, une consultation des parties prenantes a été organisée. S'appuyant sur des représentants d'Organisme Bovins Croissance (OBC) et d'Organisme de Sélection (OS), elle a permis de définir les contraintes à prendre en compte et les résultats attendus de l'algorithme de planification.

Cinq grands principes en ont été retenus concernant les caractéristiques de l'algorithme qui doit :

- Permettre une simulation prévisionnelle des dates de passages en début de campagne, basée sur les naissances des animaux de l'année précédente (N-1).
- Permettre d'ajuster les planifications en cours de campagne (pour s'adapter aux naissances réelles et aux pesées réalisées),
- Pouvoir exclure des périodes (ex : estive ou phase de mise l'herbe) et considérer les sorties anticipées des animaux,
- Pouvoir fonctionner de façon autonome sur une API (Application Programming Interface) interrogeable depuis n'importe quelle application, à partir des données minimales transmises en entrée
- Offrir la possibilité de pondérer le nombre de visites nécessaires et leur positionnement en fonction de la situation de l'élevage et de la structure.

Dès le début du projet, le développement d'une application RShiny nous est apparu nécessaire (Figure 1). Celle-ci a d'abord permis de comprendre la problématique et son contexte. Elle a ensuite été très utile pour la prise en compte des différents paramètres et contraintes exprimés par les parties prenantes. Enfin, elle a été support des tests et des démonstrations du fonctionnement de l'algorithme au fil de l'avancement de son développement.

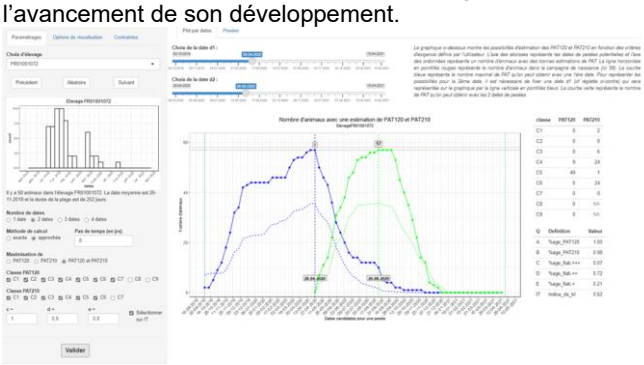

**Figure 1 :** interface RShiny

### 1.2. DEVELOPPEMENT DE L'ALGORITHME SOUS R

Le développement de l'algorithme a été réalisé par itération. Des échanges réguliers entre les équipes (développement, pilotage, utilisateurs) ont permis d'avancer pas à pas en tenant compte des besoins et des contraintes exprimés.

#### 1.2.1 Trois fichiers d'entrée sont nécessaires (Figure 2)

Le premier fichier concerne les caractéristiques de la demande de calcul. Elles contiennent les identifiants de l'élevage et l'OBC dont il dépend, la campagne et la race concernée, la date et le lot de la demande de calcul.

Les paramètres de calcul sont contenus dans le deuxième. Ils précisent le ou les PAT à optimiser (recherche de dates permettant d'obtenir le maximum de celui-ci) et le niveau minimum de fiabilité des PAT souhaité. Il contient aussi les dates de visites à exclure ou privilégier (périodes plus ou moins favorables), l'âge de sortie habituel des veaux et le nombre de visites envisagé (0 laisse l'algorithme en décider). Enfin, il permet également de paramétrer la sortie selon le niveau de détail souhaité.

Le dernier fichier d'entrée concerne les caractéristiques des animaux de l'élevage. Il contient leur date de naissance et de pesées déjà réalisées, s'ils sont à exclure de la planification et si leur poids de naissance est utilisable dans le calcul des PAT.

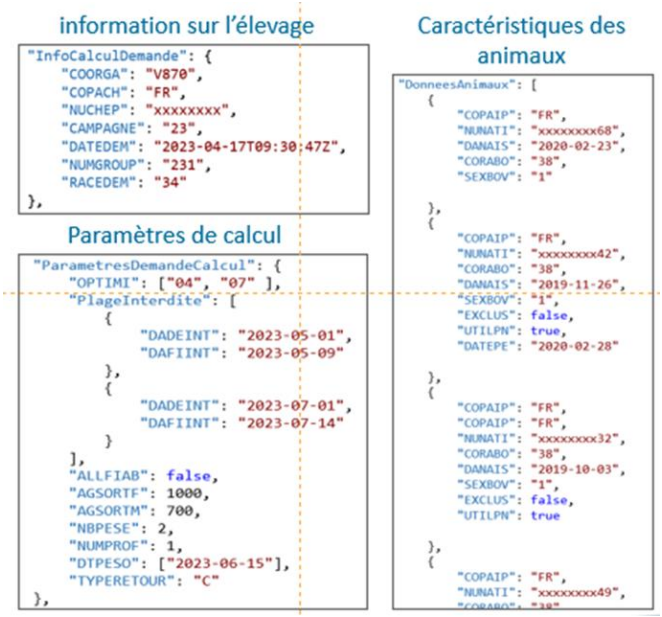


**Figure 2 :** description des fichiers d'entrée de l'algorithme

### 1.2.2 L'indice de tri et de sélection

Un indice de tri (IT) a été proposé afin d'objectiver le choix de la meilleure solution à proposer. Il combine la proportion de PAT obtenus/nombre d'animaux présents, le niveau de fiabilité de ces PAT (Delpeuch A., 2021) regroupé en 3 classes (élevé, moyen ou faible et issues de la grille détaillée contenant 7 à 9 classes selon le PAT) et le nombre de visites nécessaires. L'IT est exprimé entre 0 (faible intérêt) et 1 (fort intérêt). Deux équations sont proposées selon le souhait d'optimiser un seul PAT (P120 ou P210) ou les 2 PAT :

- Optimisation du PAT120 :

$$IT = A * (cC + \mathrm{d}D + eE) * \frac{1}{\sqrt{J}}$$

- Optimisation des deux PAT :

$$IT = (A * B) * (cC + dD + eE) * (fF + gG + hH) * \frac{1}{\sqrt{J}}$$

Avec :

- A, B : % de PAT (P120 ou P210) respectivement obtenus,
- C, F : % de ces PAT avec un niveau de fiabilité élevé,
- D, G : % de ces PAT '' '' '' moyen,
- E, H : % de ces PAT '' '' '' faible,
- c, d et e : pondérations appliquées à chaque niveau de fiabilité pour PAT120,
- f, g et h : pondérations appliquées à chaque niveau de fiabilité pour PAT210,
- J : le nombre de visites de la prévision.

### 1.2.3 Le calcul

L'algorithme dispose de deux modes de calcul :

- Exhaustif et détaillé

Dans ce mode, le calcul se fait en partant d'une grille de solutions en fonction des caractéristiques des animaux et des paramètres de la demande. Le choix de la solution se fait à partir de l'IT détaillé ci-dessus. Ce mode permet de restituer plus de détails sur la solution retenue et fournis des éléments pour ajuster les dates de façon interactive. En revanche, il nécessite un temps de calcul plus long.

- Optimisé

Ce mode permet d'aller plus rapidement à la solution adaptée aux paramètres d'entrée. Il s'appuie sur la méthode du recuit simulé (Kirkpatrick S). Cette méthode d'optimisation permet d'obtenir en un temps limité une solution optimale à un problème complexe et difficilement représentable en 2 dimensions. Elle consiste à explorer aléatoirement et de manière assez large l'univers des possibilités afin de trouver les zones les plus susceptibles de contenir la solution optimale à partir de l'IT ; puis d'affiner la recherche au sein de la meilleure de ces zones pour proposer une solution optimale. L'avantage notable de cette méthode d'optimisation est qu'elle permet d'atteindre l'optimum global (i.e. la meilleure solution) avec une très forte probabilité, à l'inverse d'autres méthodes d'optimisation qui peuvent fournir des optimums locaux (i.e. une très bonne solution mais peut-être pas la meilleure de toutes).

### 1.2.4 Evaluation des performances de l'algorithme

Pour évaluer les performances de l'algorithme, nous avons sélectionné au hasard 45 exploitations parmi les 1465 disponibles dans la base de données de test. Cet échantillon comprenait des exploitations avec 40 à 150 vêlages et un écart de naissance allant jusqu'à 10 mois (Figure 5).

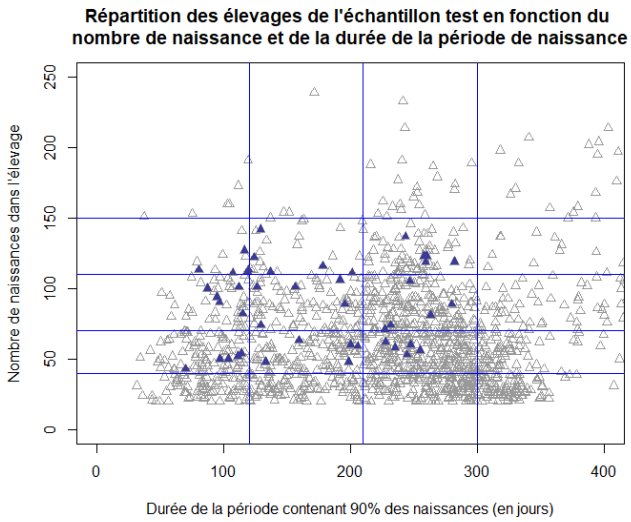


**Figure 3 :** échantillonnage des exploitations testées

Ensuite, l'algorithme a été appelé 10 fois pour chaque mode et pour chacune des exploitations sélectionnées. Nous avons ensuite calculé le temps d'exécution médian pour chaque exploitation, en ne considérant que les itérations avec le même nombre de visites.

## 2. RESULTATS

### 2.1 LES SORTIES DE L'ALGORITHME

Le résultat de l'algorithme se présente sous 3 formes différentes adaptées aux besoins et aux paramètres transmis lors de la demande. Dans le mode de calcul « optimisé » (Figure 3), la ou les dates optimales de passages par niveau d'IT sont restituées. En plus de ces éléments, le tableau de fréquence d'animaux par PAT et par niveau détaillé d'IT est fourni. Enfin, des résultats à +/- 7 jours des dates optimales sont transmis afin de permettre au technicien d'ajuster le choix final.

Dates prédites avec IT


```
"PredictionDatesPesees": {
  "ITPRED": 0.5884,
  "DATEPREDI": ["2020-04-24", "2020-09-18"]
},
```


Stats de fiabilité


```
"EffectifParPAT": [
 {
   "COAGTY": "04",
   "EffectifParNiveau": [
     {
       "NIFPAT": "C1",
       "EFFANX": 0
     },
     {
       "NIFPAT": "C2",
       "EFFANX": 0
     },
     {
       "NIFPAT": "C3",
       "EFFANX": 0
     },
     {
       "NIFPAT": "C4",
       "EFFANX": 9
     },
     {
       "NIFPAT": "C5",
       "EFFANX": 48
     },
```


Dates prédites avec IT (+/- 7 jrs)


```
"ResultatsDecales7Jours": [
 {
   "DateDecalees": {
     "ITPRED": 0.5479,
     "DATEPREDI": ["2020-04-17", "2020-09-11"]
   },
   "EffectifsDecales": [
     {
       "COAGTY": "04",
       "EffectifParNiveau": [
         {
           "NIFPAT": "C1",
           "EFFANX": 0
         },
         {
           "NIFPAT": "C2",
           "EFFANX": 0
         },
         {
           "NIFPAT": "C3",
           "EFFANX": 0
         },
         {
           "NIFPAT": "C4",
           "EFFANX": 7
```


**Figure 4 :** exemple de fichiers de sortie

Le mode exhaustif (Figure 4) fournit une sortie supplémentaire qui contient une extraction de la grille de calcul (gauche de la figure) permettant de reconstruire les courbes de nombre de PAT obtenus (continue bleue et verte) et l'IT (courbe pointillée sur une autre échelle) selon la date de visite testée.

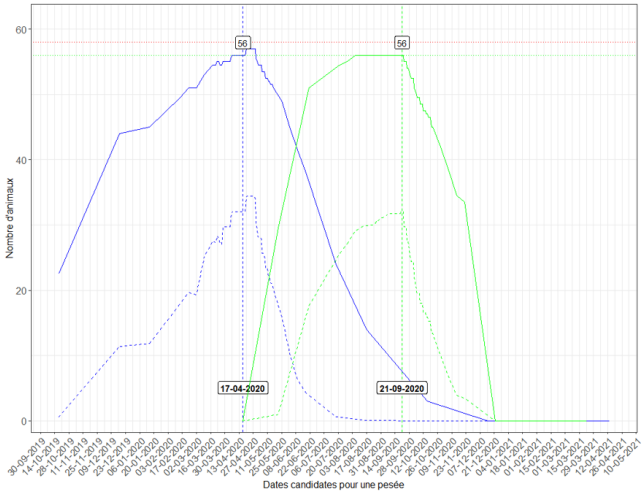


**Figure 5 :** courbes reconstruites à partir du fichier de sortie

### 2.2 LES PERFORMANCES DE L'ALGORITHME

Les performances moyennes sont satisfaisantes, surtout en mode "optimisé". Certaines fermes nécessitent un temps de calcul médian important pour obtenir des résultats (jusqu'à 9s) avec une grande plage de variabilité. Il s'agit très souvent d'élevages de grande taille avec des distributions de naissance très étalées (jusqu'à 180 jours). En revanche, le temps de calcul médian pour les élevages de taille moyenne, même avec une période de naissance étendue, est inférieur à 5s (2,4s vs 4,9s). Ces tests ont été réalisés sur un ordinateur standard et non sur un serveur dédié au calcul (Figure 6).

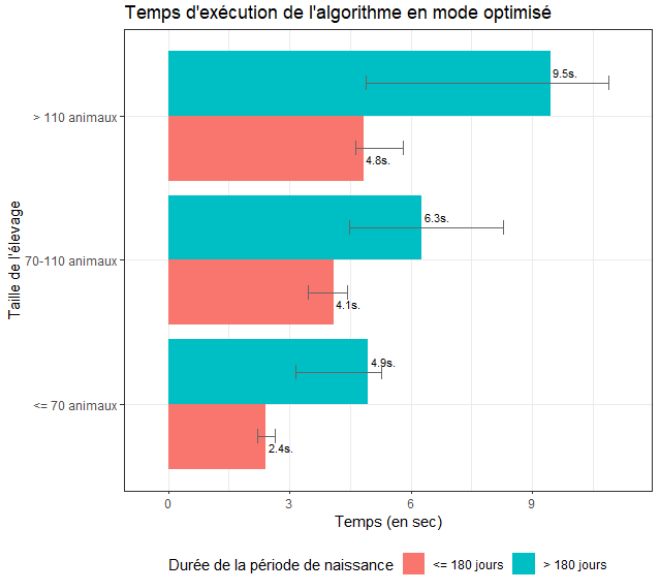


**Figure 6** temps d'exécution de l'algorithme

Pour les 21 élevages avec une période de mise-bas inférieure à 180 jours, l'algorithme fournit une solution optimale à 2 dates de visites, quel que soit le nombre d'animaux dans l'élevage.
Pour les 25 élevages avec une période de mises-bas supérieure à 180 jours, l'algorithme fournit une solution à 3 dates. Pour 3 des 5 élevages restants, l'algorithme préconise une solution à 2 dates. Pour les 2 derniers, la situation est plus contrastée avec 6 des 10 simulations avec une solution à 3 dates et 4 des 10 simulations avec une solution à 4 dates.
L'algorithme arrive à réduire le nombre de visites à effectuer dans 88,9% des cas testés. Dans 9,2% ce nombre est identique et dans 0,8% des cas, il est supérieur. Dans le cas où l'algorithme propose autant de visites que ce qui a été réellement fait dans l'élevage, l'IT augmente en moyenne de +0,27 (min = +0,05 ; med = 0,13 ; max = +0,51), ce qui témoigne d'une amélioration du positionnement de celles-ci.
Dans les cas où l'algorithme suggère moins de visites à faire que ce qui a été réellement fait, les résultats sont analogues en considérant les meilleures dates de visites réalisées dans l'élevage : l'algorithme permet d'augmenter l'IT en moyenne de +0,19 (min = -0,07 ; med = +0,12 ; max = +0,52).

## 3. DISCUSSION ET CONCLUSION

Nous avons produit un algorithme (encapsulé dans une API) qui utilise un inventaire du bétail pour déterminer les dates de pesée idéales. Cet algorithme répond aux attentes des professionnels. Sa flexibilité permet de l'adapter à un large éventail d'utilisations. Certaines options n'ont pas été implémentées dans cette première version, mais sa construction modulaire permettra son évolution dans le futur. Cette modularité dans le développement de l'algorithme permettra de faire évoluer l'outil avec les besoins, par exemple en ajoutant d'autres poids à âge-type. L'utilisation de calculs parallélisés, un affinement de la plage de recherche et une optimisation du paramétrage de l'algorithme du recuit simulé pourraient permettre d'augmenter les performances et de réduire les temps de réponse. Cela sera certainement mis en œuvre dans une deuxième version de l'algorithme.
Le service est actuellement mis en œuvre par FGE. L'API encapsulant l'algorithme devrait bientôt être mise à la disposition des utilisateurs afin que leurs logiciels puissent l'exploiter.

**Delpeuch A., 2021**, Institut de l'Elevage, "Prise en compte de la qualité des données de naissance dans le calcul des PAT et des index IBOVAL", https://idele.fr/detail-article/prise-en-compte-de-la-qualite-des-donnees-de-naissance-dans-le-calcul-des-pat-et-des-index-iboval

**Foucault M., 2023**, Institut de l'Elevage, "Résultats du contrôle de performances des bovins allaitants - France - Campagne 2022", https://idele.fr/detail-article/resultats-du-controle-de-performances-des-bovins-allaitants-france-campagne-2022

**Kirkpatrick S., Gelatt C. D. et Vecchi M. P., 1983** « Optimization by Simulated Annealing », Science, vol. 220, no 4598, p. 671–680 (ISSN 0036-8075 et 1095-9203, DOI 10.1126/science.220.4598.671)