Rédacteur du document : Institut de l’Elevage – © Tous droits réservés – 2020

GenTore\_WP5.1 : document de suivi

Développement d’un outil d’aide à la décision pour les choix des veaux à engraisser

Table des matières

[1. DESCRIPTION GENERALE 2](#_Toc45270749)

[1.1. Contexte et objectifs 2](#_Toc45270750)

[1.2. Bibliographie 4](#_Toc45270751)

[1.3. Livrables 4](#_Toc45270752)

[2. ETUDE 4](#_Toc45270753)

[2.1. Choix de la méthode 4](#_Toc45270754)

[2.1.1. Première étape : les « total merit » (TM) 4](#_Toc45270755)

[2.1.2. Deuxième étape : les « index de synthèse » 5](#_Toc45270756)

[2.1.3. Méthode de validation 5](#_Toc45270757)

[2.2. Liste des tâches 5](#_Toc45270758)

[2.3. Planification 5](#_Toc45270759)

[3. REALISATION 6](#_Toc45270760)

[3.1. Espace de travail 6](#_Toc45270761)

[3.2. Les données 6](#_Toc45270762)

[3.3. Les programmes 6](#_Toc45270763)

[4. TESTS PILOTES ET VALIDATION 9](#_Toc45270764)

[4.1. Données d’entrées 9](#_Toc45270765)

[4.2. Données de sortie 9](#_Toc45270766)

[4.3. Validation 9](#_Toc45270767)

[5. MODE D’EMPLOI 9](#_Toc45270768)

[6. MISE EN PRODUCTION 9](#_Toc45270769)

[7. COMMUNICATION 9](#_Toc45270770)

# DESCRIPTION GENERALE

## Contexte et objectifs

Ce travail est lié au WP5.1 du projet GenTore.

Le but général du projet GenTore est le suivant :

*“The objective of GenTORE is to develop innovative genome-enabled selection and management tools to optimise cattle resilience and efficiency (R&E) in widely varying environments. These tools, incorporating both genetic and non-genetic variables, will be applicable across the full range of systems (beef, milk and mixed), and will thereby increase the economic, environmental and social sustainability of European cattle meat and milk production systems”*

Les taches sont réparties de la façon suivante :



Le travail à réaliser dans le WP5 est décrit de la façon suivante :

*“Background*

*Multiple factors influence the culling risk of a cow (Berry et al., 2005), which can be summarised as the future propensity of the cow to be long-term resilient and efficient. This dictates the expected profit to be generated from the remaining lifetime of that animal. Expected future performance is influenced by many factors including the (non-)additive genetic merit of the animal for performance, accumulated environmental effects on the animal since conception, but also the genetic merit of the animal for future resilience, manifesting itself as a greater survival potential. Ideally the farmer would make culling and breeding decisions based on the knowledge of these factors integrated into an estimate of the future worth of the animal. However, tools to provide this information in an easy-to-use form are lacking. Although some progress has been made to predict culling risk from genetic information (Kelleher et al. 2016; McCarthy et al., 2007b) and also from monitoring of health and reproductive status (Richardon et al., 2014; Twomey et al., 2016; McCarthy et al., 2007a) there is a lack of tools that integrate all the above factors. Given the proposed work to better valorise automated technology for detecting resilience, including insults to the animal (e.g., lameness), and to improve the accuracy of prediction of genetic merit of an individual using genomic tools, GenTORE is well placed to tackle this gap. Accordingly, this WP offers producers the opportunity of a real-time farm environment bespoke management decision tool using high frequency phenotypic data proxies (WP3) from commercially available technologies to update predictions on the foundation of a baseline genomic index (WP4). Decision optimisation can be facilitated for breeding decisions such as sexed semen, cross breeding or culling. Results will be transferred to the relevant stakeholders in conjunction with WP7.*

*Approach*

*This multi-actor WP is desktop-based and exploits data available from both national (macro-level) and research/commercial (with detailed phenomic information from other WPs) databases for beef and dairy females as well as both phenotypic (WP3) and genomic (WP4) data generated in GenTORE. National data from Ireland (beef + dairy), the UK (beef + dairy), France (beef) and the Netherlands (dairy) will be used since beef and dairy data coexist in the same national database and these countries also represent diverse production systems; existing research data will come mainly from TEAGASC and WUR. Each dataset will be stratified into a calibration and validation dataset with the validation dataset being used to appropriately evaluate the precision and robustness of the developed approaches. Simulations will also be used for validation.*

*Phenotypic and genomic information generated in previous WPs will be used to more precisely identify mates that will produce offspring with superior lifetime potential and determine what type of breeding strategy should be used for an up-coming service (function of likelihood of pregnancy being established) based on the prevailing status of the cow as measured by real-time proxy tools (WP3). Moreover, the long-term consequence for gains of using the developed tools in the WP must be quantified to stimulate uptake by producers. Validation on a selection of commercial herds with more detailed information (RAFT) will be undertaken to increase confidence on the usefulness of the tools. Although the underlying principles will be the same for both the beef and dairy indexes, the differences that exist for the traits (e.g., availability of feed intake is available for beef and animal health data), the parameters (expected longevity differences exist between beef and dairy), the relative importance of different traits in each system, will be incorporated into the generic index framework. The impact of different strategies on environmental footprint will also be quantified for use in national carbon auditing.”*

Plus spécifiquement, la tâche 1 a pour titre « Precision ranking tool on expected lifetime female performance and resilience ».

*“Simulations generated in WP 4 will be used to generate a population of purebred and crossbred animals with genomic information and associated true breeding values and phenotypic performance. Loci-specific additive and non-additive genomic effects will be simulated. Statistical algorithms established in WP4 will be developed further to simultaneously estimate additive and non-additive genomic effects. Prediction of the expected mean (and confidence intervals) of the mating between two animals will be undertaken based on their respective genotypes and compared to the true breeding value of the simulated offspring. Mating predictions will also account for information on animals carrying causative mutations, congenital defects or variants/haplotypes with known large effects. Consideration will be given to improving the mean performance of the population but also the impact on variability in expected performance. Validation will also be undertaken using the multi-generational real-life genomic data available (WP4).This task is linked to T4.1”*

En résumé

Le but ici est de développer un outil qui permette de faire un classement intra-troupeau des veaux au moment de choisir s’ils partent à la reproduction, ou s’ils sont vendus pour l’engraissement. L’index calculé devra être une estimation de revenu généré par l’animal sur toute sa vie (coût tout au long de sa vie, prix de vente de la carcasse). Il devra donc prendre en compte tous les effets fixes et aléatoires qui peuvent avoir un impact sur la marge générée par l’animal. La valeur prédite sera exprimée à la fois en valeur monétaire et en empreinte environnementale ( ?).

Dans un premier temps, l’outil sera basé sur les caractères et effets actuellement inclus dans les évaluations nationales officielles. Dans un second temps, les prédicteurs de la résilience obtenus dans le WP3 ainsi que les résultats sur les interactions génotype/milieu pourront être ajoutés dans le model.

Pour réaliser ce travail, on se base sur les travaux déjà réalisés par les irlandais. Ils ont écrit un article (non paru encore) décrivant l’outil qu’ils ont développé. On cherche donc à adapter leur méthode aux données françaises. En France, nous allons travailler en race Charolaise et en premier lieu sur les mâles et en particulier les animaux BeefAlim (600).

## Bibliographie

* Dunne et al. (pas encore publié) : Formulation of a decision support tool incorporating both genetic and non-genetic effects to rank young growing cattle on expected market value.

## Livrables

Les livrables attendus sont un outil développé (ou au moins une méthode validée et un cahier des charges décrivant l’outil terrain à développer).

# ETUDE

## Choix de la méthode

Le principe de l’outil irlandais est d’essayer de calculer une performance pour chaque caractère à partir des résultats des évaluations génétiques et de ce qu’on connaît de l’animal au moment du tri. Ensuite, on combine ces performances à l’aide de coefficients économique afin de déterminer le revenu potentiel généré par l’animal à la fin de sa vie.

### Première étape : les « total merit » (TM)

On commence par calculer une performance potentielle pour chaque caractère qu’ils appellent « total merit ».

Ils ont 3 formules de PV possibles :

* Total Merit « Racial » : TMR= effet race seul
* Total Merit « Génétique » : TMG = effet race + hétérosis + EBV
* Production value : PV = effet race + hétérosis + EBV + effets fixes et aléatoires

Ce qu’on cherche à faire lorsqu’on calcule la PV, c’est partir de l’index (effet animal, index sur ascendance *a priori*, corrigé de tous les effets environnement) et ajouter uniquement les effets connus au moment du tri et permanents (c’est-à-dire qui peuvent avoir un impact sur le prix de la carcasse ou le coût de production). Chaque élément de la formule de PV est associé à un poids. **A priori, on pourrait utiliser les poids de l’évaluation génétique mais il est précisé dans la description du workpackage que ces poids doivent être adaptés à un objectif à plus court terme** (pas sûr que cette partie soit applicable à la situation française nous n’avons pas ce genre de poids dans l’évaluation IBOVAL).

Nous allons calculer des PV pour les caractères de carcasse (poids, conformation et âge à l’abattage), la docilité et la RFI. Nous n’avons pas d’effet race puisque les évaluations nationales sont faites en race pure uniquement :

* Nous n’aurons pas de TMR
* TMG = EBV ou GEBV ou DGV (on choisit la valeur génétique la plus précise disponible au moment du tri).

***PV des caractères de carcasse***

Les effets fixes inclus dans l’évaluation sont :

* Cellule : troupeau naisseur x troupeau engraisseur x campagne de naissance x saison de naissance
* Rang de vêlage

Le but est d’avoir un outil de classement intra-troupeau naisseur, donc l’effet troupeau naisseur est inutile, tout comme l’effet campagne de naissance (au moment du tri, les veaux seront nés lors de la même campagne)

Au moment du tri, l’animal sera encore dans son troupeau naisseur donc on n’aura pas l’information sur le troupeau engraisseur.

Finalement, les deux seuls effets qu’on connait au moment du tri et qui peuvent avoir un impact sur le revenu de la carcasse sont le rang de vêlage de la mère et la saison de naissance du veau. A noter que l’effet saison de naissance n’est pas disponible directement puisqu’il est intégré dans l’effet cellule. Donc on fait la moyenne des effets cellules correspondant à la saison 1 et la moyenne des effets cellules correspondant à la saison 2.

***PV = VG brute + effet rang de vêlage + effet saison de naissance***

**PV des caractères de comportement**

Les effets fixes inclus dans l’évaluation

* Cellules = troupeau x pointeur x sexe x conduite x lot (comprend la date de visite)
* CLAGE : classe d’âge du veau à la mesure
* PREFEM : présence des mères lors de la mesure oui/non
* DIA-SIA : distance animal-pointeur x situation (pré/box)
* SIP : situation (normal/favorisé)

Tous ces effets ont un impact sur la mesure mais n’auront pas d’impact en eux-mêmes sur la marge générée par l’animal. Par exemple, le fait que la mère était présente au moment du pointage n’a pas d’impact sur le prix de carcasse ou n’engendre pas de coût spécifique au court de la vie de l’animal.

Pour ces raisons, aucun de ces effets ne seront utilisés dans le calcul de la PV du comportement.

***PV = TMG = VG brute***

**PV de la RFI**

### Deuxième étape : les « index de synthèse »

Les irlandais calculent deux « index » de synthèse :

* Harvest index = somme des PV des caractères de carcasse pondérée avec des poids économiques associés à chaque caractère
* Calf index = harvest index + PV des caractères de docilité et d’ingestion

Nous allons donc appliquer les mêmes formules en appliquant nos propres poids économiques. L’estimation des poids économiques sera réalisée par Nathalie Malaval. Après une première analyse :

* Poids économiques des caractères carcasses : on les a
* Poids économique de la docilité : on ne l’aura pas. On ne sait pas si la docilité a réellement un impact sur la marge réalisée par l’éleveur. On va donc faire confiance aux irlandais, utiliser le poids qu’ils ont utilisé et on vérifiera lors des validations l’utilité de conserver la docilité dans la formule
* Poids économique de la RFI : pour l’instant on ne l’a pas mais on pourrait peut-être le calculer si on a la définition exacte du caractère.

### Méthode de validation

Cette partie n’est pas encore claire dans ma tête.

## Liste des tâches

La liste des tâches se trouve dans le document GenTore\_WP51 – Liste des tâches.xlsx

## Planification

Il faudrait avoir un premier test réalisé avant septembre

# REALISATION

## Espace de travail

* Répertoire des programmes {PROG} : /g2b/GENTORE/WP51/prog
* Répertoire des données {DATA}: /g2b/GENTORE/WP51/data
* Répertoire de travail {TRAVAIL} : /travail/gentore/WP51

## Les données

* Résultat des évaluations carcasses {EVALCARC} : /bdir/bovins/iboval/retour\_carcasse\_jbf
	+ Fichier perf.txt
	+ Fichier index.txt
	+ Fichier index\_geno.txt
	+ Fichier effetsfixes.txt
	+ Fichier cellules.txt
* Résultat des évaluations comportement {EVALCOMP} : /bdir/bovins/iboval/retour\_comportement
	+ Fichier perf.txt
	+ Fichier index.txt
	+ Fichier effetsfixes.txt
	+ Fichier cellules.txt
* Résultat des évaluations RFI {EVALRFI} :

## Les programmes

#### 1ers tests

**{PROG}/verif\_R/suiviRM\_WP5.Rmd**

**{PROG}/verif\_R/verif\_PV.Rmd**

**{PROG}/recup\_effets\_fixes.sas**

Le but de ce programme est de :

* Récup traitement iboval à lire et niveau des effets fixes.

Lecture de la liste des animaux beefalim, du fichier corres\_iboval, merge par station et année\_station. Transformation RVEL et SAIS en niveau comme iboval.

En sortie : {TRAV}/listeanim.sas7bdat

 4 colonnes : anim, tt\_ibo, niv\_RVEL, niv\_SAIS

* Récup des valeurs des effets fixes par traitement IBOVAL

Lecture de tous les traitements bdir disponibles (jusqu'au 2016\_01).

RVEL

Récupération de l'effet RVEL estimé à chaque traitement.

On a bien chaque niveau de l'effet qui s'exprime en + ou - x kgs de carcasse par exemple.

Création d'une table contenant pour les 3 caractères et pour chaque traitement : niveau, effectif et effet estimé.

Saison de naissance

Cet effet n'est pas directement disponible, il est en interaction dans l'effet cellules. L'effet cellule n'a pas l'air d'être exprimé par rapport à un niveau de référence.

Pour retrouver un effet saison : calcul de la moyenne des effets cellules par saison, puis la saison 1 est fixée à 0, la saison 2 est exprimée par rapport à la saison 1.

Exemple pour AGEABA :

moyenne saison 1 = 552 jours, moyenne saison 2 = 570.

Effet saison 1 = 0, effet saison 2 = +18 jours

En sortie : effet\_rvel\_car au format sas et format csv.

effet\_saisn\_carc en 2 formats sas et csv.

 dispo actuellement sur {DATA}/test\_V1 et {TRAVAIL}/test\_V1

* Fusion des informations valeurs et niveaux pour les animaux beefalim

En sortie : table sas **fictot\_effets\_carc.sas7bdat.**

Remarque : que carcasses uniquement pour l’instant car pas prévu de récupérer d’effets fixes pour le comportement.

Question : on va donc potentiellement récupérer ensuite la valeur d’un effet à un traitement où peut-être le veau n’aura pas été évalué ?

**{PROG}/recup\_index.sas**

Le but de ce programme est de créer un fichier global contenant tous les index d’une liste de traitements pour les animaux d’une liste et leurs parents. En pratique, il va :

* récupérer les parents d’une liste d’animaux
	+ Extraction animrace la plus proche de l’indexation
	+ Si la généa de la bdir est différente de celle fournie, l’animal est supprimé (on considère que la généalogie fournie est la bonne, si celle de la bdir est fausse, alors l’index est faux également)
	+ Si la date de naissance est inconnue en bdir, l’animal est également supprimé
* récupérer les index CARC et COMP, poly et geno pour l’ensemble de ces animaux (ceux de la liste et leurs parents) pour une liste de traitements
	+ Normalement, le traitement choisi devrait dépendre de l’année de naissance (pour se mettre dans les conditions réelles, on ne doit prendre que les informations connues au sevrage)
	+ La liste des numéros de traitement est à fournir en paramètres
* Calculer pour la liste d’animaux de départ, un index sur ascendance.
* Déterminer l’index le plus précis :
	+ 1. GEBV
	+ 2. EBV
	+ 3. Ascendance
* Fichiers de sortie :
	+ {TRAVAIL}/test\_{nom\_test}/global\_index\_{tt} (avec tt=numero de traitement): contient pour la liste d’animaux, leur généalogie, les index récupérés et les index calculés
		- Anim : identifiant animal
		- Pere : père (animrace.txt)
		- Mere : mère (animrace.txt)
		- Sexbov : sexe (animrace.txt)
		- Corabo : code race (animrace.txt)
		- Datenai : date de naissance (animrace.txt)
		- Annenai : année de naissance (animrace.txt)
		- VG\_poly : valeur génétique brute (index.txt)
		- DG\_poly : valeur génétique en déviation (index.txt)
		- Ind\_poly : index poly (index.txt)
		- VG\_geno : valeur génomique brute (index\_geno.txt)
		- DG\_geno : valeur génomique en déviation (index\_geno.txt)
		- Ind\_geno : index geno (index\_geno.txt)
		- VG\_asc : valeur ascendance brute (calculée à partir de la VG la plus précise des parents)
		- DG\_asc : valeur ascendance en déviation (calculée à partir de la DG la plus précise des parents)
		- Ind\_asc : index ascendance (calculée à partir de l’index le plus précis des parents)
		- VG : valeur génétique brute la plus précise (geno>poly>asc)
		- DG : valeur génétique en déviation la plus (geno>poly>asc)
		- ind : index le plus precis (geno>poly>asc)
		- VG\_cod : code origine de VG (« geno », « poly » ou « asc »)
		- DG\_cod : code origine de DG (« geno », « poly » ou « asc »)
		- ind\_cod : code origine de ind (« geno », « poly » ou « asc »)
		- VG\_per : valeur génétique brute la plus précise du père (en priorité geno, poly sinon)
		- DG\_per : valeur génétique en déviation la plus précise du père (en priorité geno, poly sinon)
		- ind\_per : index le plus precis du père (en priorité geno, poly sinon)
		- VGc\_per : code origine de VG\_per (« geno » ou « poly »)
		- DGc\_per : code origine de DG\_per (« geno » ou « poly »)
		- indc\_per : code origine de ind\_per (« geno » ou « poly »)
		- VGp\_per : valeur génétique brute du père (index.txt)
		- DGp\_per : valeur génétique en déviation du père (index.txt)
		- indp\_per : index poly du père (index.txt)
		- VGg\_per : valeur génomique brute du père (index\_geno.txt)
		- DGg\_per : valeur génomique en déviation du père (index\_geno.txt)
		- indg\_per : index geno du père (index\_geno.txt)
		- VG\_mer : valeur génétique brute la plus précise de la mère (en priorité geno, poly sinon)
		- DG\_mer : valeur génétique en déviation la plus précise de la mère (en priorité geno, poly sinon)
		- ind\_mer : index le plus precis de la mère (en priorité geno, poly sinon)
		- VGc\_mer : code origine de VG\_mer (« geno » ou « poly »)
		- DGc\_mer : code origine de DG\_mer (« geno » ou « poly »)
		- indc\_mer : code origine de ind\_mer (« geno » ou « poly »)
		- VGp\_mer : valeur génétique brute de la mère (index.txt)
		- DGp\_mer : valeur génétique en déviation de la mère (index.txt)
		- indp\_mer : index poly de la mère (index.txt)
		- VGg\_mer : valeur génomique brute de la mère (index\_geno.txt)
		- DGg\_mer : valeur génomique en déviation de la mère (index\_geno.txt)
		- indg\_mer : index geno de la mère (index\_geno.txt)

On ne trouve pas tous les index COMP et CARC poly et geno pour l’ensemble des traitements de la liste puisque ces évaluations n’ont pas été mises en production à la même date. Je n’ai pas automatisé cette partie. Je fais donc tourner le programme en plusieurs fois en

* modifiant la liste des traitements récupérés
* en mettant en commentaire le type d’index manquant.
* Modifiant le nom du fichier de sortie

**{PROG}/ fusion\_index.sas**

Ce programme fusionne les fichiers « global\_index » produits par le programme « recup\_index ». On obtient un seul fichier global\_index.

# TESTS PILOTES ET VALIDATION

## Données d’entrées

**Liste des animaux beefalim :**

Elle se trouve sous {DATA}/liste\_beefalim/liste\_beefalim.csv.

La description du format du fichier se trouve sous {PROG}/bao/file\_formats/liste\_beefalim.sas

**Le pédigrée des animaux beefalim :**

Répertoire source {PEDIG} : /bdir/bovins/commun/donnees\_bn/ex\_{extr}/r38

Pour le test {extr} = 2020\_06\_26

**Résultats d’indexation :**

Répertoire source {BDIR} : /bdir/bovins/iboval

Pour le test, nous avons utilisé les index :

* Carcasse JBF poly cha :
	+ {BDIR}/retour\_carcasse\_jbf/poly/{tt}/cha
	+ Avec {tt} = du traitement 2014\_02 au traitement 2020\_02
* Carcasse JBF geno cha :
	+ {BDIR}/retour\_carcasse\_jbf/geno/{tt}/cha
	+ Avec {tt} = du traitement 2016\_01 au traitement 2020\_02
* Comportement poly cha :
	+ {BDIR}/retour\_comportement/poly/{tt}/cha
	+ Avec {tt} = du traitement 2016\_02 au traitement 2020\_02

## Données de sortie

**Le fichier global\_index :**

Comme on ne trouve pas tous les index COMP et CARC poly et geno pour tous les traitements de 2014\_02 et 2020\_02. J’ai donc fait tourner le programme recup\_index 3 fois et j’ai obtenu en sortie 3 fichiers :

* {TRAVAIL}/test\_V2/global\_index\_2016\_02 : contient tous les types d’index du traitement 2016\_02 au 2020\_02 inclus
* {TRAVAIL}/test\_V2/global\_index\_2016\_01 : contient tous les index CARC du traitement 2016\_01
* {TRAVAIL}/test\_V2/global\_index\_2014\_02 : contient les index CARC poly du traitement 2014\_02 au 2015\_02 inclus

J’ai fusionné ces trois fichiers avec le programme fusion\_index pour obtenir un fichier unique :

* {TRAVAIL}/test\_V2/global\_index
* Je l’ai également copié sous {DATA}/test\_V2

## Validation

# MODE D’EMPLOI

# MISE EN PRODUCTION

# COMMUNICATION