

MASTER PROJECT

A Web-GIS application
for the monitoring of
Farm Animal Genetic Resources
(FAnGR)
in Switzerland

June 15, 2014

Author:
Solange Duruz

Supervisor:
Stéphane Joost
Expert:
Christine Flury

Environmental Science and Engineering section
Laboratory of Geographic Information System (LaSIG)

Abstract

In 2007, FAO (Food and Agriculture Organization, from the United Nations) initiated the Global plan of action for Farm Animal Genetic Resources (FAnGR) to reduce further loss of genetic diversity in farm animals. One of the key issues mentioned is to identify endangered breeds to support conservation prioritization programs.

In this context, the Swiss Federal Office for Agriculture attributed a mandate to explore the feasibility of the implementation of a monitoring concept. The report mentioned the relevance of including the geographic location of the populations monitored. Accordingly, we used open source software (PostgreSQL, PostGIS, OpenLayers, Geoserver), to develop a Web-GIS platform prototype (GenMon-CH) designed to assess pedigree information, geographical concentration, socio-economic and environmental information.

GenMon-CH includes PopRep developed by the Institute of Farm Animal Genetics (FLI, Germany) to run the pedigree analysis and to provide parameters such as inbreeding coefficient, effective population size. Additionally introgression will be considered. Current developments will soon make it possible to process these indices based on genetic information as well. In parallel, the combined socio-economic/environmental index assesses the attractiveness and the risk of potential future agricultural practice abandonment in the regions where populations are bred. Finally, a multi-criteria decision support tool aggregates criteria using the MACBETH method, which is based on a weighted average using satisfaction thresholds.

The system permits to upload basic information for each animal (parents, birth date, sex, location, introgression) and to choose relevant weighting parameters and thresholds. Based on these inputs, the system completes a pedigree analysis, and computes a final ranking of breeds based on an integrated conservation prioritization score to be visualized on a map.

Résumé

En 2007, la FAO (Food and Agriculture Organization des Nations Unies) a initié le “Global plan of action for Farm Animal Genetic Resources (FAnGR)” destiné à réduire les pertes potentielles de diversité génétique chez les animaux d'élevage. Un des problèmes-clés est l'identification des races menacées dans le cadre de programmes de conservation par priorités.

Dans ce contexte, l'Office Fédéral de l'Agriculture en Suisse a accordé un mandat pour explorer la faisabilité de l'implémentation d'un concept de monitoring. Le rapport mentionne la pertinence d'inclure la position géographique des populations surveillées. Sur cette base, nous avons développé un prototype de plate-forme Web-GIS (GenMon-CH) en utilisant des logiciels open sources (PostgreSQL, PostGIS, OpenLayers, Geoserver), conçue pour traiter l'information sur le pedigree des animaux, la concentration géographique des populations ainsi que des facteurs socio-économiques et environnementaux.

GenMon-CH comprend le logiciel PopRep développé à l' "Institute of Farm Animal Genetics" (FLI, Allemagne) pour effectuer l'analyse de pedigree et pour estimer des paramètres tels que le coefficient de co-sanquinité et la taille effective de la population. De plus, le degré d'introgession est également pris en compte. Les développements en cours rendront bientôt possible le calcul de ces indices à partir de données génétiques. En parallèle, un indice socio-économique et environnemental évalue l'attractivité des régions où les races sont élevées et le risque d'un potentiel abandon des pratiques agricoles dans ces dernières. Un outil d'aide à la décision multi-critères agrège les variables en utilisant la méthode MACBETH, basée sur une moyenne pondérée utilisant des seuils de satisfaction.

Le système permet de charger des informations de base pour chaque animal (parents, date de naissance, sexe, lieu, introgession) et de choisir les paramètres d'agrégation adéquats (poids et seuils de satisfaction). Sur la base de ces données, le système effectue l'analyse du pedigree et calcule un rang final pour la race en question, basé sur un score intégré correspondant au risque d'extinction qui peut être visualisé sur une carte.

Contents

1	Introduction	5
2	Method	6
2.1	Building a global index	6
2.1.1	Pedigree Analysis	7
2.1.2	Introgression	8
2.1.3	Geographical concentration	8
2.1.4	Local- and Breed-Agriculture Sustainability	9
2.1.5	Multi-criteria aggregation	11
2.2	Detailed information of breeds under investigation	13
3	Data	14
3.1	Animal and breeds	14
3.2	Socio-Economic and landuse variables	15
3.3	Geographic units	16
4	Prototype implementation	17
5	Results based on 2 case studies	20
5.1	Summary table	20
5.2	Franches-Montagnes	20
5.2.1	Inbreeding	20
5.2.2	Effective population size	21
5.2.3	Maps	22
5.3	Original Braunvieh	24
5.3.1	Inbreeding	24
5.3.2	Effective Population Size	25
5.3.3	Map	25
5.4	Local Agriculture Sustainability index	27
6	Discussion	27
6.1	Comments on the result	27
6.2	Pros and cons of the method	28
6.3	Technology chosen	29
7	Conclusion	30
	References	33

8 Appendices	35
A Thresholds applied in different countries	35
A.1 Food and Agriculture Organization (FAO)	35
A.2 United Kingdom	36
A.3 Germany	36
B Software used in pedigree analysis	36
C Important parameters used in pedigree analysis	37
D Thresholds and weights used	39
E Details of the land use scenario of the WSL	40
F Install GenMon-CH on a new server	42

Note that a **glossary** is available at the end of the document.
Unless otherwise specified, italic words correspond to glossary entries.

1 Introduction

Today, the livestock sector is facing an important stress due to the increased food supply demand. This has led to the selection of more productive breeds which in turn generated a loss in the genetic diversity of farm animals (Tisdell, 2003). Nevertheless, genetic diversity is essential to face future challenges including climate change (Notter, 1999). Indeed, it contributes to an increased resistance of species to diseases (Chee-Sanford et al., 2001).

Based on these observations, the *Food and Agriculture Organization (FAO)* proposed a “Global Plan of Action for Animal Genetic Resources” presented at the Interlaken conference in 2007. The countries involved in that plan must implement policies, identify threats to genetic diversity and build monitoring systems. (FAO, 2007a)

In Switzerland, the *Federal Office for Agriculture (BLW)* is responsible for this monitoring. It attributed a mandate to the Bern *School of Agriculture, Forest and Food Science (HAFI)* to explore the feasibility of the implementation of a monitoring concept. The report mentioned the relevance of including the geographic location of the populations monitored, and recommended to build the monitoring system and framework. Together with the *Ecole Polytechnique Fédérale de Lausanne (EPFL)* they are implementing this system.

One of the major problems however, is to define the criteria that are meaningful in order to identify breeds that are endangered. The *FAO* created a scale of endangerment based on the number of breeding females. Thus, a breed is declared not at risk if “the total number of breeding females and males is greater than 1000 and 20, respectively; or if the population size is greater than 1200 and the overall population size is increasing” (FAO, 2007b). This definition has the advantage of being simple but is an oversimplification of the problem. Indeed, other criteria need to be taken into account for this purpose, and the decision might depend on the size of the country, the type of landscape and the species considered (Alderson, 2010).

Currently, three countries have already set up a monitoring system of cattle diversity: Austria, Germany and Great Britain. Austria mainly uses the number of female animals while Germany and Great Britain also include genetic or *pedigree* information. They assess this latter aspect by first computing the *inbreeding (F)* coefficient, which estimates the relatedness of ancestors. This parameter is then used to determine the *effective population size*. Great Britain includes the geographical concentration of the breeds as well, since a breed that is clustered in a small region is more vulnerable to epidemics (Alderson, 2010). (DEFRA, 2013; BLEMV, n.d.)

The *inbreeding (F)* coefficient is in fact often used in the literature regarding genetic diversity. There are two approaches to get this parameter. The first one is to use DNA-analysis while the second one estimates it from the *pedigree* data (Keller & Waller, 2002). In Switzerland, DNA-sampling in livestock is not performed on a regular basis; as a result, the second alternative is the only option that can currently be applied. Many people have already considered the problem of *pedigree* analysis (see Groeneveld et al., 2009, for example) and several software solutions already exist that can possibly be used for this purpose (see the description in appendix B).

Besides the above-mentioned factors, other parameters can also be found in the literature related to breed endangerment. For instance, Alderson (2010) specifies that *introgression* should be considered since the trait of a specific breed will be mixed by this process.

Socio-economic factors can also be relevant for the monitoring of endangered breeds. Even though this kind of criteria is less frequently found in the literature, Alderson (2009) defines a few socio-economic factors that are pertinent, such as the age of the breeders and the likelihood of catastrophe. While being pertinent criteria, the feasibility of including such aspects still has to be assessed. For example, the age of the breeder can be known when a survey is performed within the study (see Bertaglia, Joost, & Roosen, 2007, for example). However, if this is not the case, it is difficult to know the age of the breeder. Likewise, the likelihood for catastrophe to happen is hard to define. Other approaches have been proposed, for example by using socio-economic as well as environmental data of a region (e.g. demographic balance, percentage of agriculture and landuse) (Bertaglia et al., 2007; Joost et al., 2010). The idea here is to assess the sustainability of breeding activities (and agriculture in general) of a given region.

This long list of criteria points out the question of how to consider all these different aspects. Here we propose to apply *Multi-Criteria Decision Analysis (MCDA)*-techniques. A wide diversity of aggregation methods exist, some of which are easy to implement but might be a little too simplistic, while other certainly better represent the complexity of the reality but are more difficult to implement. A given method must then be chosen, that best fits the problem. (Guitouni & Martel, 1998)

Furthermore another relevant question arises, which is to know how to characterize the animal populations with socio-economic variables available at the regional level. The link is in fact achieved through the geographical information and the spatial coincidence concept (Goodchild, 1996). Thus, by knowing the location of an animal, it is possible to connect it with socio-economic statistics of a region. (see for example, Joost et al., 2010)

All these elements are taken into account in the present project, whose goal is to develop an easy-to-use Web-*GIS*-application to be used for decision support, allowing to identify the breeds that are endangered and the regions that need help. It includes the following criteria: *pedigree* analysis (from *herdbooks*), introgression, geographical concentration and agriculture sustainability (processed on the basis of socio-economic and environmental data). The application generates a ranking of the most endangered breeds based on these criteria using an *MCDA*-approach. It also allows the user to quickly identify the prominent problems of a given breed. Furthermore the regions that show problems are highlighted, according to a specific criterion (e.g. high *inbreeding (F)* in a given region) and the importance of that region for the breed is visible as well (in terms of number of animals).

2 Method

2.1 Building a global index

The above-mentioned ranking of the breeds will be performed based on a global sustainability index, which includes 4 aspects, that are in fact sub-indices. For the sake of comprehension, each sub-index will be abbreviated, and the following nomenclature will be used throughout the rest of the paper. To better understand how these indices are used, refer to figure 1.

- *Pedigree* analysis index: **pedig-index**

- *Introgression* index: **introg-index**
- Geographical concentration index: **geog-index**
- Local agriculture sustainability index: **LAS-index**
- Breed agriculture sustainability index: **BAS-index**
- Global sustainability index: **global-index**

The *LAS*-index is the aggregation at a regional level, independently from the presence of animals in the territory. The *BAS*-index in turn is the average of the *LAS*-index over the region in which a breed is present. The link between those two is achieved through geographical location.

The pedig-, introg-, geog- and *BAS*-indices are aggregated to constitute the global-index. The overall procedure is given in the following figure, while the detailed description of each index is available in the upcoming sections.

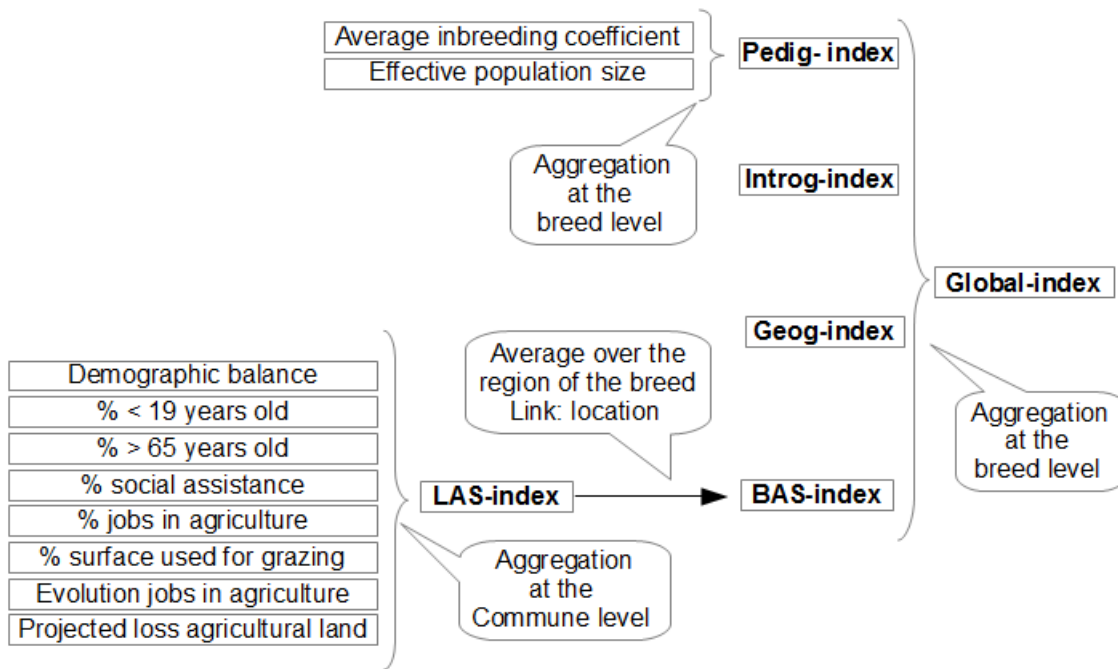


Figure 1: Criteria aggregation

2.1.1 Pedigree Analysis

For the pedig-index, there are two parameters that are often mentioned in the literature: the *inbreeding* (F) coefficient and the *effective population size* (N_e). These two parameters are thus taken into account in this project.

Regarding the first parameter, the average over one year (i.e. the last year for which animal are present the data) could be used; however, if the data of the last year are incomplete, the global vision of the breed will be erroneous. This is why it has been decided to use the average *inbreeding* (F) coefficient over the last *generation interval*

(*GI*), instead of over the last year. On the other hand, this choice has the disadvantage of reducing the relative importance of recent data and to reveal a greater inertia if a sudden change occurs.

Regarding the *effective population size* (N_e), several methods exist to compute this parameter and they all provide very different results (Gutiérrez et al., 2008, see also appendix C). Therefore, it is not easy to choose the most appropriate procedure for a specific case. In PopRep (the software used for the *pedigree* analysis; see section 4), N_e -indices are summarized in a module called PR-Mon (Groeneveld et al., 2009). This module also selects the best-appropriate method to compute N_e based on the time-frame used by each technique. Based on previous experiences, it has been noticed that for some breeds, this module can output unexpected results and is not well-adapted for the case of Switzerland (Christine Flury, personal communication, 09.04.2014). One aspect that came out of this discussion is that the N_e parameter does not need to be very precise, and that a range of values is sufficient. For these reasons, PR-Mon is not used in the application; instead, the user is being presented with the different values of N_e calculated by the different methods and he must choose a range. The N_e values are also put in comparison with the *pedigree* completeness for the method based on the *inbreeding* (F) change of rate. The range chosen by the user will be stored in the database and will be used to calculate the pedig-index. If in the future PR-Mon is further developed and the outputs are found to be satisfactory, this module could be included in the GenMon-Ch pipeline.

2.1.2 Introgression

GenMon-CH is designed to directly store the *introgression* rate calculated elsewhere. The question remains however to know which animal should be taken into account to calculate the mean *introgression*. For example, for breeds that are largely introgressed, ancestors from one or several other breed(s) should be present in the *pedigree* data. In this case, their *introgression* is of 100%. Should these animals appear in the computation of the mean *introgression*? This is an open question; nevertheless, for the moment, they are taken into account. Another litigious example is the case where animals from another breed that do not (yet) have offsprings are in the reference population and thus appear in the *pedigree* file.

Instead of entering directly the *introgression* rate in the system, the breed of the dams and sires could be specified in the inputs as an additional attribute. In this case the *introgression* could be calculated on this basis. However, this option is not implemented in GenMon-CH for the moment.

To build the introg-index, the average over the last *generation interval* is computed. The time-frame of one *generation interval* has been chosen for the same reasons as the frame chosen for *inbreeding* (F) coefficient.

2.1.3 Geographical concentration

The geographical concentration is calculated as the smallest circle containing at least 75% of the animals, centred around the centroid of the breed. The centroid is computed as the weighted mean coordinates of regions containing animals. Given that a precise position of animals is not available, the centroid of *ZIP-code*-polygons containing animals is used as an approximation.

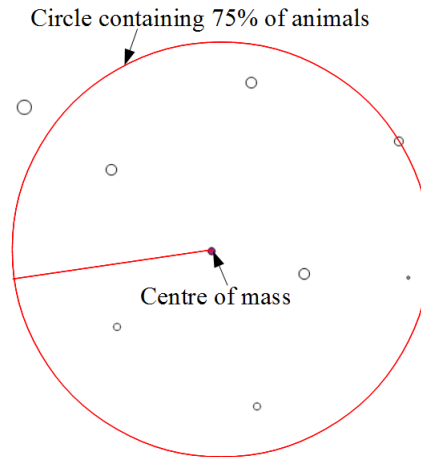


Figure 2: Concept to quantify geographical concentration

The geog-index thus corresponds to a radius (in km).

2.1.4 Local- and Breed-Agriculture Sustainability

The *LAS*-index is based on socio-economic and environmental factors in a given region (i.e. the region where a breed is located). Indeed, a breed that is located in regions where agricultural practice are being abandoned can be considered as endangered (Bertaglia et al., 2007) even if the *pedigree* analysis of that same breed reveals no problem. A similar use of socio-economic data is also mentioned in e.g. Joost et al. (2010). The goal of this index is to assess the capacity of a region to sustainably host breeding activities. This is achieved by first performing a global analysis of the dynamic of the region, and by then looking specifically at the agriculture and breeding activities in that same area. This procedure is completed at the commune-level since it is the level at which socio-economic data are available.

First, the overall attractiveness of the region is being assessed. This includes the demographic balance between a year in the past and now (the *Federal Statistical Office (BFS)* computes directly this variable based on a 2-years difference; see section 9), which shows if people are rather leaving or coming to the commune. Next the age population structure is considered, which comprises two variables: the percentage of people younger than 19 years and older than 65 years respectively. Indeed a commune that is ageing will face problems in the future if nothing happens (e.g. financing of the pensions, of the hospital cares, ...). Alderson (2009) proposes to use the age of the breeder. This of course would be a better alternative; however this parameter is difficult to evaluate and the corresponding data are not available. Finally, the percentage of social assistance is taken into account. The initial idea was to use the unemployment rate; however this parameter is not available at the commune level in the official federal statistics. The idea behind this parameter is that if the social assistance rate is high, people will have the tendency to go away to find a job to find a more attractive and dynamic place to live.

Then, a series of parameters assesses the importance of agriculture in general and of grazing in particular in a commune. If a commune is mainly rural, it will probably be more aware of the problems that farmers are facing and might also be willing to make

efforts to support this sector. In this context, the parameters used are the percentage of farmers and the percentage of grazing surfaces. The recent evolution of agricultural practice is assessed as well. This includes the balance of jobs in the primary sector between a previous year and now (see more information in section 9).

Last but not least, a layer predicting the future land use elaborated by the *Swiss Federal Institute for Forest, Snow and Landscape Research (WSL)* is used (Price et al., 2013). In this layer, the land use for 2050 is computed. By comparing it to the present time, we can assess if agricultural land are likely to be gained or lost over the next few decades. Mainly two kinds of losses are considered in these scenarios: agricultural land abandonment with subsequent forest growing as well as urban sprawl. The scenarios also take into account the consequences of the predicted climate change. The computation has been done according to five scenarios that are described in further details in appendix E. They are based on different policies as in the IPCC scenarios (Houghton, 2000) with added policies and hypotheses. With the intent of facilitating the use of GenMon-CH, only one scenario is taken into account in this application. Since we do not know what the policies will look like in the future, it has been decided to use the trend scenario. It seems to be a reasonable outcome, since unless very strict policies are setted up, the trend that has been noticed over the last few decades is likely to continue.

Once the *LAS*-index is computed at the commune level, the geographical link with the breed is made to get the *BAS*-index (at the breed level). Note that statistics are available for a commune but the animals are located via their *ZIP-code*. Furthermore, the *ZIP-code*-polygons do not necessarily match the commune-polygon. The *LAS*-index of a *ZIP-code* is computed as the *LAS*-index of the commune in which the centroid of the *ZIP-code*-polygon lies. Note that individuals constituting a breed are distributed among several regions. The *BAS*-index is thus computed as an average of the *LAS*-index on regions where the breed is bred (weighted by the number of animals per *ZIP-code*).

The two following figures are here to better understand these concepts. The first one illustrates how the average over the territory is done, by taking into account only areas that contain animals (in blue), while other (in white) do not appear in the computation. The second one shows the concept of spatial coincidence.

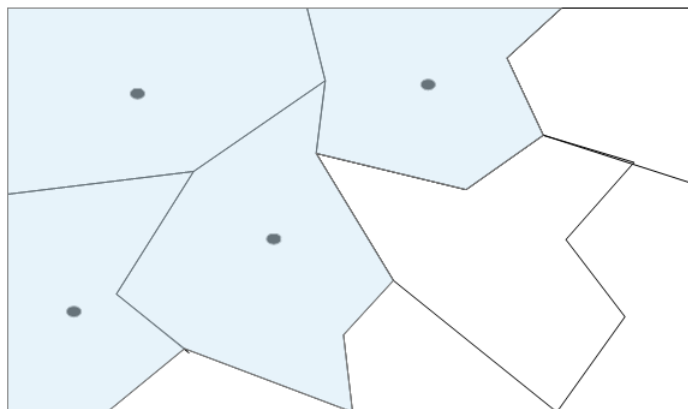


Figure 3: Average over areas that contain animals

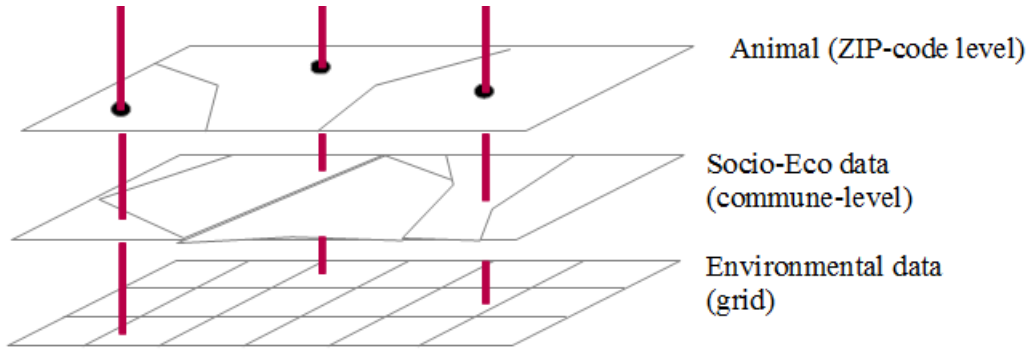


Figure 4: Concept of spatial coincidence

2.1.5 Multi-criteria aggregation

In this case, the method *MACBETH* (Costa et al., 1994) was chosen which is a compromise between simple weighted average and a more complex method named *ELECTRE* (Roy, 1991). See section 6.2 to know the reasons that have led to this choice.

For each criterion, the user must decide a weight (to distinguish the criteria that are the most important) and thresholds values, below and respectively above which the criterion is not at all respectively completely satisfactory. Using the thresholds, the initial values of a criterion is thus rescaled (normalized) to give a satisfaction score whose values are between 0 and 1 (or 0 and 100%).

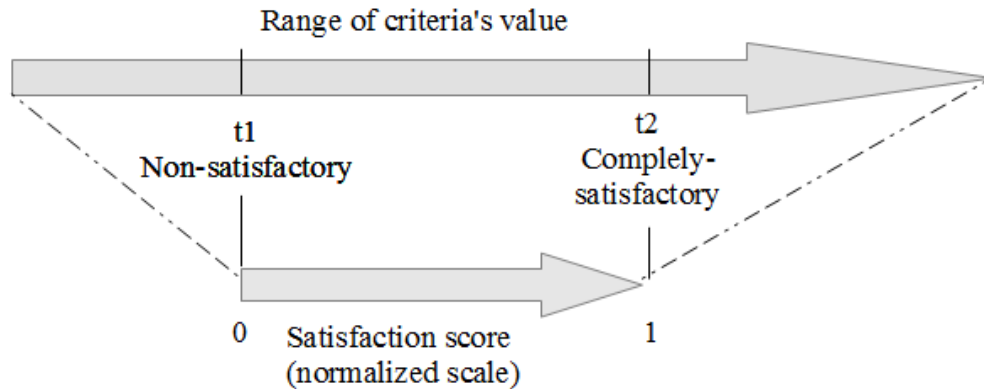


Figure 5: Normalization using the MACBETH method

Then criteria are aggregated using the following formula (weighted average of the satisfaction scores):

$$S = \sum_{j=1}^J s_j * w_j \quad (1)$$

Where:

j: the criterion considered (J being the total number of criteria)

s: the satisfaction score

w: the weight (while the sum of the weights equals 1)

The satisfaction score is computed using the following idea

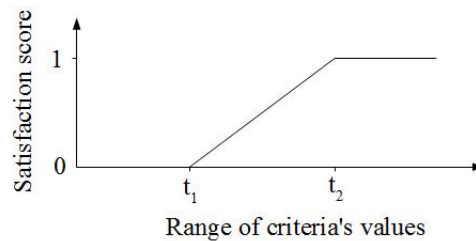


Figure 6: Satisfaction score with the MACBETH method

Where

t_1 : the non-satisfaction threshold (defined by the user). Below this threshold the score obtained by the criteria is the lowest (i.e. 0% satisfaction).

t_2 : the total-satisfaction threshold (defined by the user). Above this threshold the score obtained by the criteria is the highest (i.e. 1 or 100% satisfaction). (Costa et al., 1994)

The following example illustrates the concept. To build the global-index, the thresholds for the geog-index must be chosen. As a reminder, this criterion corresponds to the radius of the circle containing 75% of the population. Based on the scale used in the U.K. it can be decided to choose $t_1=12$ (km) and $t_2=50$ (km), thus giving the following figure:

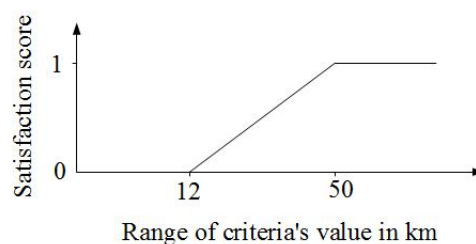


Figure 7: Example of satisfaction score for the geog-index

If the breed has a geographic concentration radius of 5 km (initial value = 5) it will have the same score (satisfaction score = 0) as a breed with a radius of 12 km. On the other hand a breed with a radius of 80 km is considered as well spread, the same way a breed with a geog-index of 50 km is (satisfaction score = 1). Scores for values between 12 and 50 km are obtained via the tilted line. For example a radius of 31 km would result in a score of 0.5.

Pedig-index		Introq-index		Geog-index		BAS-index		Global-index
score	weight	score	weight	score	weight	score	weight	
0.6	0.5	0.8	0.2	0.5	0.1	0.9	0.2	0.69
0.3	0.5	0.9	0.2	0.1	0.1	0.7	0.2	0.48

Table 1: Example of the computation of the global-index

This methodology is applied to all criteria included in the considered index. The final score is the average, weighted by the weights of each criterion. The following table gives an example of the computation of the global score for two hypothetical breeds.

As a reminder, the *MACBETH*-method is used at different levels: 1) to compute the sub-indices (for example the pedigree-index) containing itself several criteria (i.e. *inbreeding (F)*, *effective population size* in that case) and 2) the global-index. The detailed description of the procedure was shown in figure 1.

2.2 Detailed information of breeds under investigation

The method described above is used to create a global-index using criteria aggregation. Once this is achieved, it is important to be able to go into further details and identify the issues of a given breed and the regions where these problems occur.

As a result, it is possible for each breed to access detailed information on the *pedigree* analysis either by looking into the reports (generated by PopRep, see section 4) or by looking at the more succinct version of this document provided by GenMon-CH. Indeed, the most important tables and figures of the reports are directly displayed by the application.

Besides this function, the interface is also conceived to allow the visualisation of the spatial distribution of the breeds and of some other parameters. Maps are designed to identify the regions that show problems and to determine which criterion is the most problematic (e.g. is it a *pedigree* problem or a socio-economic problem). As a result, the user can visualize the average *inbreeding (F)* coefficient and *introgression* rate over the last *generation interval* per *ZIP-code* as well as the maximum *inbreeding (F)* coefficient. Therewith, the number of individuals in these *ZIP-code*-polygon can be visualized. Furthermore, the *LAS*-index can be visualized on the map too. All this information is given by a choropleth map that shows the class into which a given *ZIP-code*-polygon falls; 5 layers can be selected and displayed and the interface allows the user to switch from one layer to another. Moreover, it is possible to know the exact value for the above-mentioned parameters by clicking in the target polygon. The map-service is also designed to provide statistics over a wider region; in this case, several polygons can be selected simultaneously and GenMon-CH displays the average or total value.

3 Data

3.1 Animal and breeds

GenMon-CH is designed to accept data described below. The standard format is specified in the parenthesis, but the user also has the opportunity to go through an “assisted upload” section in which he can specify the format of the data.

- a) Animal ID (any series of characters without special characters, e.g. é)
- b) Sire ID (any series of characters without special characters, e.g. é)
- c) Dam ID (any series of characters without special characters, e.g. é)
- d) Year of birth (e.g. 2009. The user can specify other formats e.g. DD/MM/YYYY)
- e) Sex (M/F or as specified by the user)
- f) *ZIP-code* (e.g. 3096)
- g) *Introgression* rate (between 0 and 1 or if specified 0 and 100. The user can also enter the *introgression* as a fraction of its own blood instead of foreign blood, if he specifies so).
- h) *Inbreeding (F)* coefficient calculated with genetic tests (between 0 and 1 or if specified 0 and 100)
- i) Cryo-conservation (1 if the gametes of the animal is being cryo-conserved, 0 if not)

In the standard format, columns are separated by a semicolon (;) and the first line is intended for headers. This can be changed in the “assisted upload” section. An example of file in the standard format is given here:

```
animal_ID;sire_ID;dam_id;birth;sex;plz; intro;inb;cryo_cons
73400;70335;358651;1951;M;3057;0;;0
398242;16590;7756462;1994;F;1971;0.1;;0
```

Figure 8: Example of data in their standard format

The first five parameters (a to e) are required to run the *pedigree* analysis (any animal with missing information in these fields will be deleted from the database). The other parameters can be left empty if data are not available and will be simply ignored during the computation. If the *ZIP-code* is missing, the geog-index will not be computed and it will not be taken into account in the computation of the final index. Furthermore, the mapping-service will not be available. Note also that for the moment, the last two fields are not used; they will be used in the context of developments of the prototype.

For the moment, two breeds have been entered in the database: the Franches-Montagnes (FM) horses and the Original Braunvieh (BVO) cattle. The data were provided by breeding organizations and entailed the parents, birth, sex, *ZIP-code* and *introgression* rate.

Introgression and *ZIP-code* were not available for all animals, but only for the most recent ones.

For the Franches-Montagnes, the *pedigree* tree included 42982 animals, born between 1854 and 2013, from which 26496 were born during the last 10 years.

For the Braunvieh, data from 84031 animals born between 1900 and 2014 were used. Out of these animals 37331, animals were born during the last 10 years.

3.2 Socio-Economic and landuse variables

These variable can be obtained from the Federal Statistical Office^{1 2}. They should be updated as often as possible (approximately once a year when statistical data from the previous year are released). They include the following (the text in italics corresponds to the exact name used by the *BFS*):

- j) Increase/decrease in population in the last 2 years (in %); *Residents - change in %*
- k) Social assistance rate (in %); *social assistance rate*
- l) Number of jobs in the primary sector; *Employed total - primary sector*
- m) Total number of jobs (all three sectors); *Employed total*
Out of these two parameters, the percentage of jobs in the primary sector will be computed
- n) Surface used for animal breeding (in ha) *grünfläche* from STAT-TAB
- o) The total surface of the commune (in km²); *Total surface area in km²* from STAT-TAB
Out of these two parameters, the percentage of grazing surface will be computed
- p) The number of jobs in the primary sector from a previous year (the year should be decided by the user, and the corresponding threshold set accordingly); *Employed total - primary sector* from STAT-TAB. This allows to compute the evolution of jobs in the primary sector considering two years
- q) Percentage of the population younger than 19 years old (in %); *Age distribution in % - 0-19 years*
- r) Percentage of the population older than 65 years old (in %); *Age distribution in % - 65 years or over*

Data used for the case study were extracted from the “regional portraits 2014: Communes” (BFS, 2014), except for variables l) and o) that came from STAT-TAB 2012 as well as p) from STAT-TAB 2010.

Besides the data uploaded by the user, landuse *GIS*-layers from the *WSL* are also used to compute the *LAS*-index (Price et al., 2013). In their original format, they consist of six

¹Regional portraits - communes: <http://www.bfs.admin.ch/bfs/portal/en/index/regionen/02/key.html>

²STAT-TAB Landwirtschaftliche Betriebe nach Jahr und Gemeinde: <http://www.pxweb.bfs.admin.ch/Dialog/statfile.asp?lang=1prod=07>

files (current state and five scenarios) containing ascii grids with cells of 1 ha resolution (see appendix E). Only the layers with the “trend” scenario and the current state have been used here, and a comparison between these two states has been carried out: the cells within *ZIP-code*-areas that currently constitute agricultural land (arable or meadow) and are predicted to remain in this same category in the future according to the chosen scenario are counted. The percentage per *ZIP-code* of these non-changing cells with respect to the number of agricultural cells at the beginning is then computed. Thus, the percentage represents the proportion of the agricultural land that will still be used for agriculture in the future according to the projections.

3.3 Geographic units

The geographical units allow to link the different dataset (animal, socio-economic factors) on the basis of their geographical location (spatial coincidence). The first file includes the *ZIP-code*-shapefile from the June 2013 as defined by the Swiss Post, which allows to approximately georeference an animal according to its *ZIP-code*. It is a shapefile containing 4191 polygons, which represent the spatial extent associated with a given postal code. Note that it is possible to have two distinct polygons with the same *ZIP-code*. Therefore, if two polygons had the same *ZIP-code*, the two geometries have been merged into multipolygons in order to make the *ZIP-code* unique (primary key in the database) and make the subsequent computations faster. The file is in the swiss coordinate system (CH-1903). The polygons have been simplified (using *QGIS* simplify geometry function with a precision of 20m) to reduce the size of the file and thus the time needed to load the data on a map.

Likewise, the shapefile of the communes in 2013 is also used (Swisstopo, 2014), which is required to georeference the *LAS*-index of a commune. The link between the socio-economic data and their corresponding polygon is made through the BFS number (unique ID of the commune, used by the *BFS*). The file attribute are the BFS number and the geometry. It is in the swiss coordinate system (CH-1903) and contains 2564 polygons.

The following figure summarizes the types of data that are used and shows how they can be linked through geography so that in the end, everything is referenced at the *ZIP-code*-level.

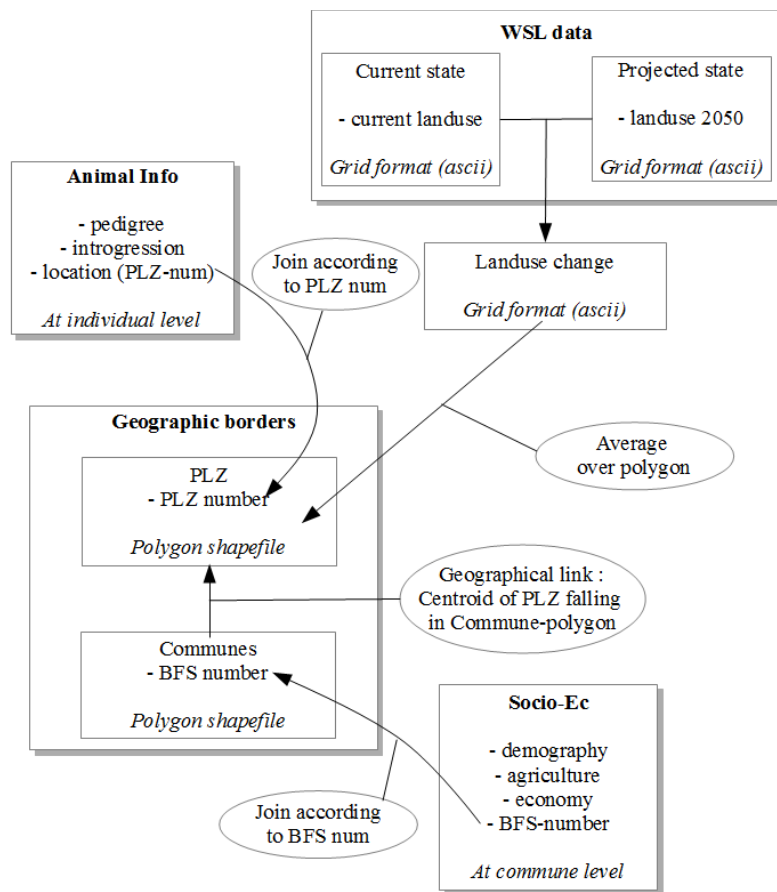


Figure 9: Data types and their link

4 Prototype implementation

Several options currently exist to build a Web-*GIS* interface. One of the requirements that had to be complied with was that the software on which it is built must be open source. Here is a description of the main parts of the technology that has been used. The reasons that have led to the choice of the solutions described in this section is further developed in the discussion.

The main part of the interface is built in *Hypertext Markup Language (HTML)* and *Hypertext Preprocessor (PHP)* language. The upload of the file (containing either animal information or socio-economic variables) is made through an *HTML*-form. The file is then stored in a *database management system (DBMS)*. For this project PostgreSQL³ is being used with its spatial extension PostGIS.

Once stored in the database, some queries are sent in order to prepare the data for further computation (e.g. change into strictly numerical ID) and the *pedigree* analysis is completed by PopRep (Groeneveld et al., 2009). PopRep is an open-source software coded in perl often used in *pedigree* analysis. It produces three output reports. The first one describes the population structure (number of males/females per year, *generation interval*...), the second one gives information about the *inbreeding (F)* and the third one (called

³www.postgresql.org

PR-Mon) computes the *effective population size* of a breed. As already mentioned this third report is not used. Before printing the reports, data are stored in a PostgreSQL database. The outputs used from PopRep are mainly the *generation interval* (from population report), the *inbreeding* (F) coefficient and the *effective population size* (inbreeding report). This allows to compute the *inbreeding* (F) coefficient over the last *generation interval*. The PopRep code is run with the *PHP* command “exec”.

Then, *Structured Query Language* (*SQL*)-queries are sent to aggregate values: averages and sums over the last *generation interval* for the whole breed as well as per *ZIP-code*. The indices described before are computed. The weights and thresholds used for this computation must be provided by the user before the upload and are stored in the database. PopRep most important tables are stored before all other temporary tables are deleted.

For the visualisation part, the interface is built on OpenLayers⁴. Openlayers is an open-source javascript library, which offers a great flexibility. The map is generated with a Web Mapping Service (*WMS*) through the use of Geoserver⁵. Geoserver is an open source software written in java that allows to edit geospatial data. Given that the shapefile is large (more than 4000 polygons with multiple edges), a *Web Mapping Service* (*WMS*) rather than a *Web Feature Service* (*WFS*) seemed appropriate to map the geometries. As a result, the map is shown through *WMS* as an image. This map is inserted in the *HTML*-code by adding a javascript code using openlayers library. The map can show the attributes of the polygons (for example the mean *inbreeding* (F) of the commune) through the use of *Extensible Markup Language* (*XML*). In the *XML* file, the attributes values are classified (the values of the class limits must be defined manually). For a given class, the background color of the polygons can also be defined. Given that the map is displayed as an image, the *WMS* is not considered dynamic; the only option is zooming. As a result, it does not allow to get information on a given object other than the corresponding class of the specified attributes (given by the color of the polygon).

To fetch more information, a *WFS* request must be built (*WFS* `getFeature` in OpenLayers). This fetches the geometry as well as the attributes of a given polygon (in that case the polygon that is selected). This also allows to select multiple polygons and to show statistics on them (as for example the total number of animals in the selected zone or the mean *inbreeding* (F)). The statistics are calculated using javascript language.

Geoserver does not allow a dynamic change of the number of fields of the table to be mapped. Indeed, once a new column is added to this table, the feature must be manually reloaded. Thus, it is not possible to store the information of all breeds per *ZIP-code* in this table since the number of breeds to monitor is likely to change over time. To overcome this problem, a new table per breed is created with the results of the aggregation per *ZIP-code*. When the mapping environment is called, a join with the table containing the spatial information of the *ZIP-code* is done. To increase the performance of this computation, primary and foreign keys have been created in these tables. The system keeps track of the last breed that is stored in the spatial table through a log-table. If the mapping-service is required for a breed, the system will first check in the log if the breed is stored in the spatial table; if so, it avoids the join-step. This system has the disadvantage of complicating the process of comparing the spatial distribution between two or more breeds.

⁴www.openlayers.org

⁵www.geoserver.org

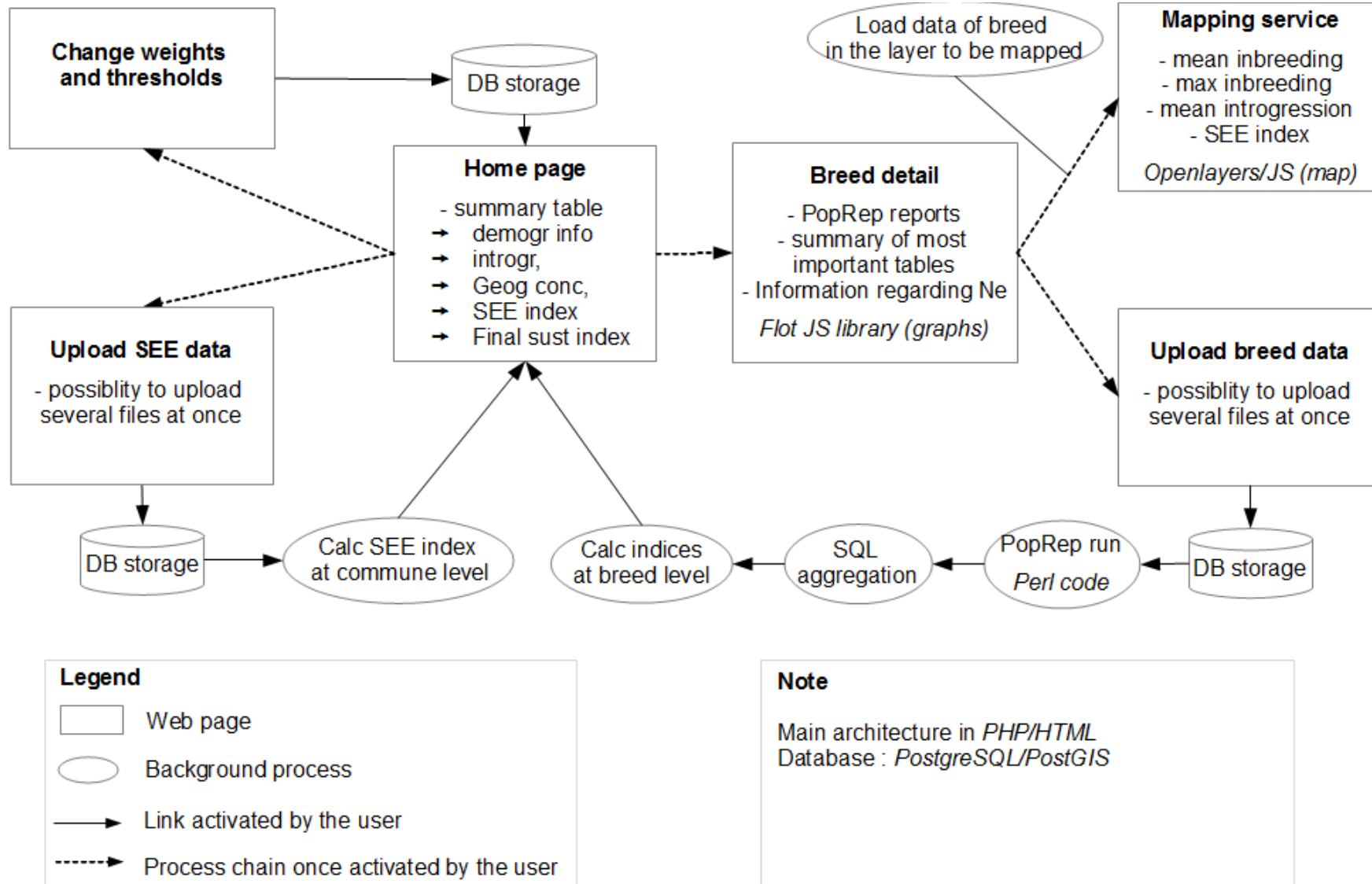


Figure 10: Scheme of the GenMon-CH application web-flow

5 Results based on 2 case studies

Two breeds have been entered into the database for the present research: the *Franches-Montagnes (FM)* horse and the *Original Braunvieh (BVO)* cattle. Data from other breeds are supposed to complete the database in the future.

5.1 Summary table

One of the key-outputs of GenMon-CH is a summary table in which breeds are ranked according to their global-index.

Breed name	Num animals (last GI)	Mean F (last GI)	N_e range	pedig-index	introg-index	geog-index	BAS-index	global-index
FM	26877	0.057	50-100	0.383	0.114	57.66	0.75	0.454
BVO	30470	0.033	50-100	0.594	0.013	58.52	0.74	0.812
XXX	...							
XXX	...							

Table 2: Summary table

For an explanation of the names of the indices, please refer to section 2.1.

By comparing the two breeds, we first see that the *FM* shows a much lower global-index than the *BVO* does. As a reminder, the global-index is included between 0 and 1. It also has an average *inbreeding* (F) coefficient that is slightly higher than the *BVO* one, but above all, it has an *introgression* that is way higher than the *BVO*.

5.2 Franches-Montagnes

5.2.1 Inbreeding

The following figure represents the average and maximum *inbreeding* (F) over that last 20 years.

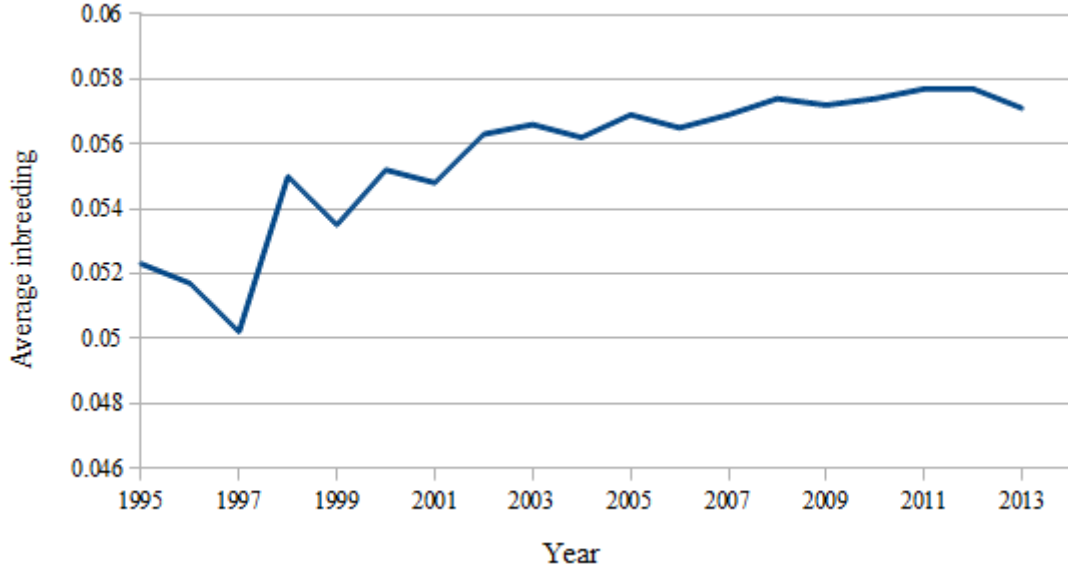


Figure 11: Mean and max inbreeding over the last 20 years for the Franches-Montagnes

The mean *inbreeding* (F) is not very stable at the beginning; however, it should be noted that the range of variation is of 0.05, which is very small. Then, around the year 2002, it gets steadier, reaching a value of around 0.058. The slope since that year is very gentle.

5.2.2 Effective population size

This table gives a summary of the *effective population size*, computed with different methods. For more details on the computation, see appendix C.

Method	N_e
$N_{e\Delta Fg}$	101
$N_{e\Delta fg}$	120
N_{eln}	100
$N_{e\Delta Fp}$	73

Table 3: N_e for the Franches-Montagnes

Note that the method $N_{e\Delta Fp}$ gives good estimate of the *effective population size* only if the pedigree completeness is good. In this case, the pedigree completeness to the sixth generation approaches 100% (i.e. ancestors of the animals born within the last 20 years are known at least to the sixth generation).

5.2.3 Maps

Now that possible problems have been identified, it is important to know where these issues are particularly visible.

The first map shows the average *inbreeding* (F) coefficient over the last *generation interval* (in this case 10 years) per *ZIP-code*.

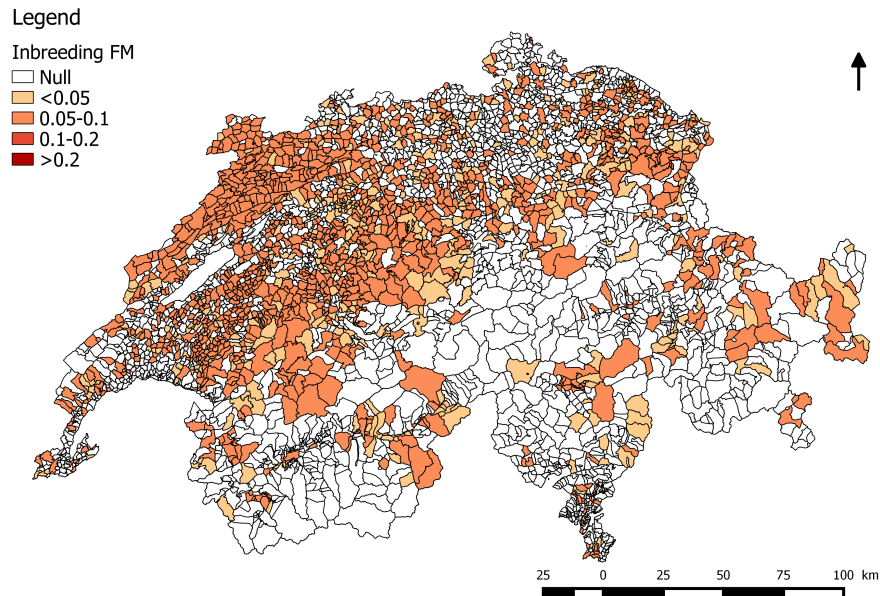


Figure 12: Mean inbreeding over the last generation interval for the Franches-Montagnes

The distribution of this coefficient is fairly homogeneous. There is no region with extreme values.

Then it is possible to have a look at the *introgression* coefficient.

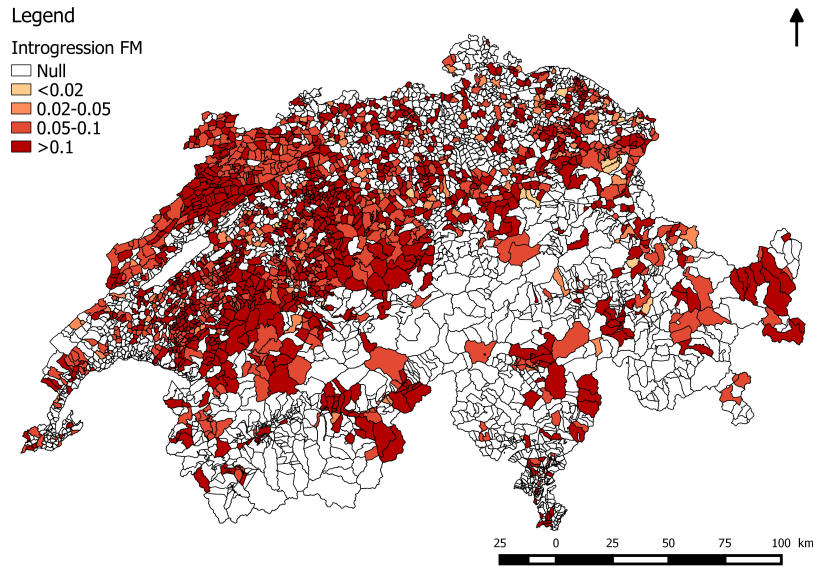


Figure 13: Mean introgression over the last generation interval for the Franches-Montagnes

As already pointed out in the summary table the Franches-Montagnes breed has high *introgression* values. Therefore, it is not surprising to see high values over the whole country. There are only few regions with low *introgression* rate. We can therefore conclude that the *introgression* problem of this breed is not a regional problem but instead is a global issue.

It is always good to put the previous results with the relative importance for the breed of the previously identified regions. Here is a map of the number of individuals per *ZIP-code*. This map only makes sense when compared with the previous ones, since absolute numbers should not be directly displayed on a choropleth map.

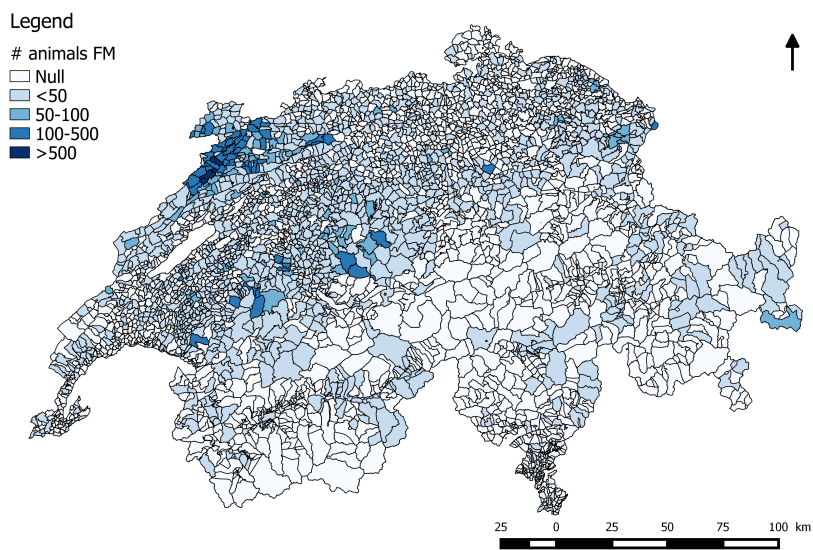


Figure 14: Number of animals over the last generation interval for the Franches-Montagnes

The breed is mainly concentrated in the Jura-region (from where the breed originates). By comparing the three maps, we can identify that the regions with the most individuals have, like almost all other regions, an acceptable *inbreeding* (F) coefficient but a high *introgression* rate.

The GenMon-CH application would actually allow a better assessment of particular regions than paper maps, since it is possible to click on the map to obtain the precise value of the parameters (*inbreeding*, *introgression*, ...) assigned to a *ZIP-code* or group of *ZIP-code*. It is also possible to zoom in the map when the user is interested by one particular region, and to change layer dynamically.

5.3 Original Braunvieh

5.3.1 Inbreeding

The evolution of the *inbreeding* (F) over the last 20 years for the breed *BVO* is presented in the following graph.

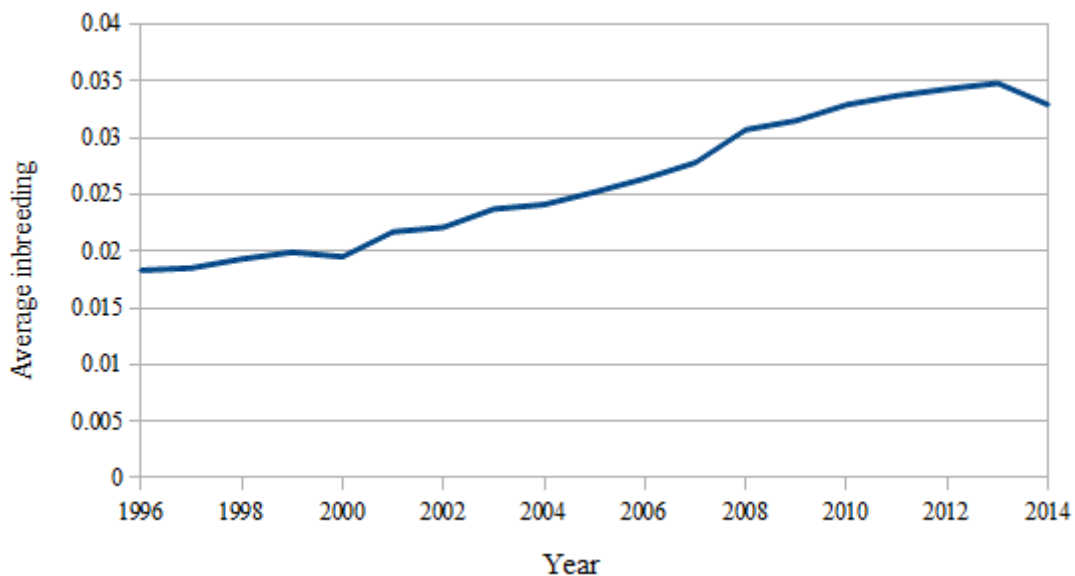


Figure 15: Mean and max inbreeding over the last 20 years for the Original Braunvieh

Here, the average *inbreeding* (F) is more stable than for the *FM* breed, given that the slope is more regular. The slope is a bit steeper, but average values are very low.

5.3.2 Effective Population Size

Method	Ne
$N_{e\Delta Fg}$	-2
$N_{e\Delta fg}$	-2
N_{eln}	111
$N_{e\Delta Fp}$	85

Table 4: N_e for the Original Braunvieh

Note that the *effective population size* cannot be negative. The negative results given in the table are actually due to computational problems and should not be taken into account.

Again, for the $N_{e\Delta Fp}$, the pedigree completeness should be assessed to know the accuracy of that method. In this case, the pedigree completeness almost reaches 100% to the sixth generation.

5.3.3 Map

The distribution of animals and parameter is presented here, by first displaying the map of the *inbreeding* (F) coefficient over the last *generation interval* (7 years for the case of *BVO*).

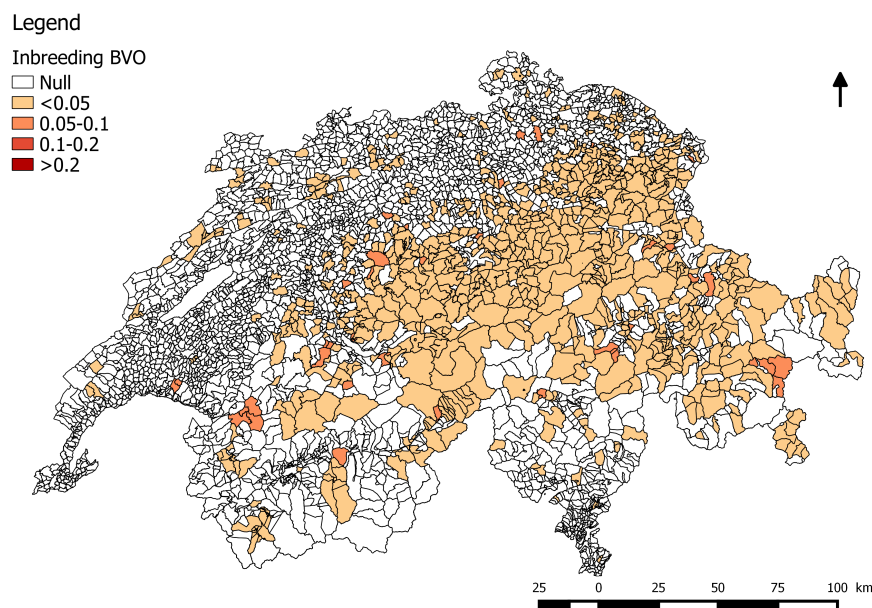


Figure 16: Mean inbreeding over the last generation interval for the Original Braunvieh

Here this coefficient is quite low everywhere, even though some ZIP-code with a bit higher values are noticeable.

Regarding the *introgression*, the results can be seen in the following map.

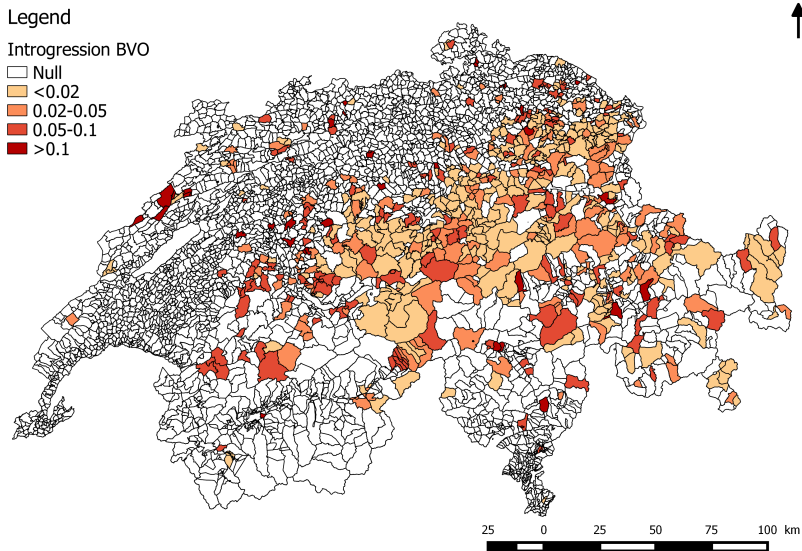


Figure 17: Mean introgression over the last generation interval for the Original Braunvieh

Here, the *introgression* is not homogeneous at all. Overall values tend to be quite low; nonetheless, some regions have very high values. This is the case in the West, in the region called the “Vallée de Joux”.

Again the previous maps should be compared with the one showing the number of individuals.

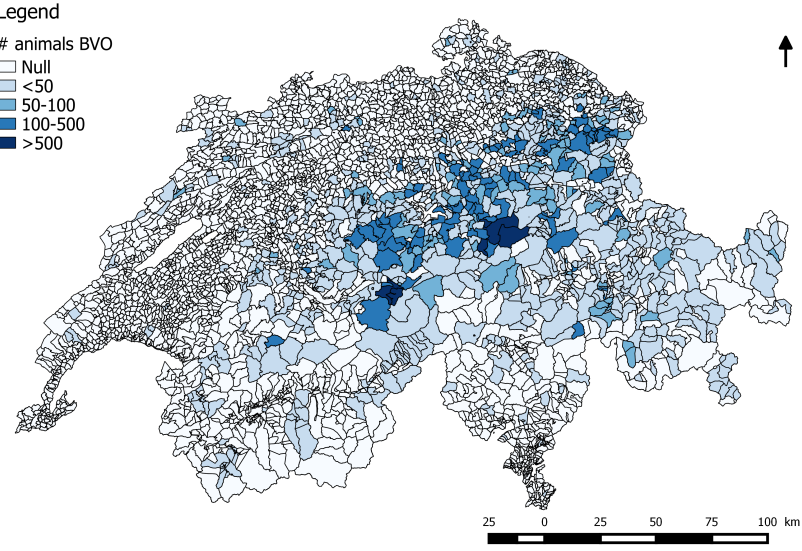


Figure 18: Number of animals over the last generation interval for the Original Braunvieh

Most of the animals are located in the Swiss-German Pre-Alps. It is relevant to see that the *ZIP-code* with many animals have low values of *introgression* and *inbreeding* (F). On the other hand, the “Vallée de Joux” for example have only a few animals and the issue caused by that region having a high *introgression* rate is thus minimized.

5.4 Local Agriculture Sustainability index

Here is a map of the *LAS*-index (that is used for both breeds).

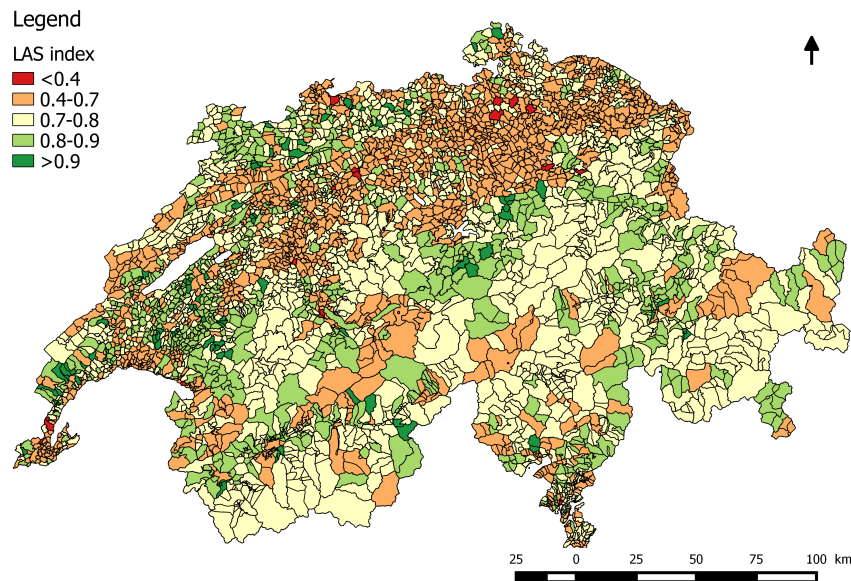


Figure 19: Local Agriculture Sustainability index (2014)

The lowest values are found on the Plateau, which is where most cities are located. Beside the clear difference between urban and rural areas, we can still identify changes within rural regions. Indeed, outside of cities, some regions are dark-green (0.9-1) while other are yellowish (0.7-0.8) and even orange (0.4-0.7). Even though the range of values for rural areas is not very wide (most of the rural areas are between 0.6 and 0.8), differences among regions are still visible.

6 Discussion

6.1 Comments on the result

By looking at the summary table, it can be shown that the *FM* breed is more endangered than the *BVO*, because the former has a smaller global-index than the latter. The numbers in the summary table can be compared to threshold values used in other countries (see appendix A) to have a point of comparison. The main problem for the *FM* horse is the *introgression*. Even though this parameter cannot be compared to values used in other countries, values in the literature indicate that a mean *introgression* of more than 10%

is already quite high (Alderson, 2010). Apart from that, it has a mean *inbreeding* (F) that is acceptable (even though it is a bit higher than the one from the *BVO*) and a good geographical spread.

The *BVO* in turn is a sane breed. It has been entered in the database not because it is a priori endangered but because it is of international interest. The global-index is good and it shows no particular problem.

Regarding the evolution of the *inbreeding* (F) coefficient, it is reassuring to see that the increase in the average *inbreeding* (F) coefficient follows a gentle slope.

By looking at the *effective population size*, it can be seen that this parameter is quite low compared to threshold values in other countries (see appendix A). For both breeds it is around 100. Some methods give a bit higher score, but given that the pedigree completeness is good, the $N_{e\Delta Fp}$ can be trusted to give accurate result (Groeneveld et al., 2009). In the U.K. for example, such a low estimate of N_e would result in the classification of the breed as critical. However, it has been shown that the *BVO* breed is a priori not endangered. Therefore, it is critical to put this number in perspective with the situation of the country. Indeed, Switzerland is a small country, and it could be an explanation for such low *effective population size*. As a result, the choice of threshold values for the N_e -parameter will have to be made independently from the choice made by other countries.

With respect to the *BAS*-index, both breeds have similar values, which does not allow a memorization of the breeds based on that index.

6.2 Pros and cons of the method

As pointed out by the results shown in the previous section, the application GenMon-CH allows a good identification of endangered breeds (with the ranking of the breeds). Moreover, a rapid identification of their problems is possible (with the detailed view of each breed). This could greatly help people who are expected to allocate funds.

Furthermore, this application has the advantage of taking into account a wide variety of criteria. With the use of geographical information, it allows to link various types of data (i.e. *pedigree* information as well as socio-economic factors).

The chosen approach also allows some flexibility, since the user can decide the thresholds and weights that he thinks are meaningful in the context he is working on. If he believes, for example, that a variable does not make sense, he can ignore it by putting its weight to zero. Furthermore, depending on the species and the regions that are considered, weighting parameters might need to be set accordingly. Moreover, the satisfaction thresholds used in the *MACBETH* method are important because they work in a similar way than the scales used in other countries (see appendix A). Indeed, these scales define values below which the breed is considered as endangered and values above which the breed is not at risk and make steps in-between. Here the steps do not directly exist but instead are represented by a regular line. This *MCDA*-technique thus seems to be appropriate.

However, it can be argued that the final ranking of the breeds is based on one single number that is difficult to interpret because computed from different criteria. This approach could be designated as technocratic. While it is true that the final result is a single number, it should be reminded that the interface also allows to go into further details to analyse where the problems are visible.

Additionally, it should be noticed that the weights and thresholds that are taken into account can significantly affect the results. Given that this interface is meant to be used to help addressing subsidies for endangered breeds, it is possible that a user with some interest in a specific breed will set the weights in order to help that same breed. Indeed, if the user has an a priori idea of the problems of the considered breed, he can set the weights of the criterion that causes problem very high so that this breed will appear more endangered than what it really is. This problem does exist and it is important to keep it in mind; nevertheless, if the interface is used correctly, this issue should be minimized.

Furthermore, the computations that are made are intensive (especially the *pedigree* analysis), which requires time. Nevertheless, it should be reminded that all computations are on server-side, thus allowing the user to do something else while waiting for the results. The time greatly depends on the number of animals, but as an example, when uploading a typical pedigree of around 100'000 animals, the required time is of around one day. On the other hand, computations for a pedigree with around 1000 individuals will take only about one minute. In the future, some of the calculations that are not considered necessary can be avoided. In fact, the most time-consuming step is the part of the report on "Additive Genetic Relationships", which might be excluded from the report if needed. If this part is dropped, the resulting computation time for 100'000 animals is lessened to a few hours.

6.3 Technology chosen

As mentioned previously, one important constraint is that the technologies chosen must be open source.

The core of the interface is coded in *HTML-PHP*. In fact, most of the Web-services are using these languages allowing a good graphical output (*HTML* and *Cascading Style Sheet (CSS)* language) while having the possibility to have an object-oriented language with *PHP*.

The *DBMS*, PostgreSQL has been chosen since it is one of the most efficient open source *DBMS*. It also easily communicates with an interface built in *PHP* and with *QGIS*. This last point might be an advantage if some special *GIS* operations need to be done on the data (that are not implemented in the interface). Moreover, the PopRep code used for the *pedigree* analysis is built on a PostgreSQL database, which facilitates the data transfer.

Regarding the PopRep analysis, some alternatives do exist (see appendix B). One of the requirements for the software to be used is that it must be supported by *Unix operating system (OS)*, which is the *OS* of the server on which GenMon-CH runs. Not all software offer this advantage. Furthermore, PopRep has already been successfully used by the *BLW*, and several people there are familiar with its outputs. PopRep also has the advantage of directly creating detailed reports which can be useful.

Regarding the mapping environment, plenty of ways to show a map on the web exist. Some are easier than other, but usually, the simpler it gets, the less flexibility is available. For example, when creating a map using *QGIS*, it is possible to publish the map made in quite easily using a plugin, for example QGISCloud⁶. It is also possible to use independent solutions as for example GeoDjango⁷. However, when using QGISCloud, it is not always

⁶www.qgiscloud.com

⁷www.djangoproject.com

easy to link the rest of the interface with the plugin. The solution with Openlayers in turns offers a large flexibility (almost everything is possible) but the coding is not always straightforward and can be time-consuming. This is probably the biggest disadvantage of this technology. Nevertheless, with a bit of experience, this difficulty can be overcome. Furthermore, Openlayers is often used when dealing with Web-*GIS*, and thus is well documented and well functioning.

The decision to use Geoserver is not an important choice, and it could easily be switched to another map-publishing server if required.

7 Conclusion

The GenMon-CH application appeared within the scope of the “Global Plan of Action for Farm Animal Genetic Resources” launched by the *FAO*. It is an easy-to-use Web-*GIS* application relying on open source software solutions that provides a multi-criteria approach for the monitoring of endangered breeds. By using the geographical information to link data, the application allows to use different types of criteria, including *pedigree* data, *introgression*, socio-economic and environmental aspects and geographical concentration. It gives a simple score for each breed, allowing for a rapid identification of the most endangered breeds, while still giving the opportunity to identify the most important problems and their geographical location. The study of two breeds (Franches-Montagne horse *FM* and Original Braunvieh cattle *BVO*) showed this concept. It has thus been possible to see that *BVO* has no particular problem; it has low *inbreeding* (*F*) and *introgression*, it is spread throughout a large portion of the territory and is bred in regions that are agriculturally sustainable. The *FM* horse in turn has an *introgression* problem, which lowers its final score.

The application is currently usable, and some breeds have already been entered. However, some work still needs to be done, to have it used at a larger scale. Some of the tasks are related to the internet network. A few examples are: security issues, abortion of the script when errors occur, work queue if several users are uploading data at the same time. To some extent, this is already done but might be improved. Then the application needs to be tested by specialists (e.g. potential users) who can give their feedback on what needs to be changed, so that it fulfils at best their expectations. Moreover, depending on the evolution of the livestock monitoring, new approaches might be chosen. A relevant example is the potential future use of genetic data instead of pedigree information to avoid the time-consuming pedigree analysis step. The upload of genetic data is already possible but is not used in further computations.

Glossary

effective population size The effective population size (N_e) represents the “number of individuals in an ideal population that would have the same genetic response to random processes as a real population of size N ” (Ellstrand & Elam, 1993). It is usually computed based on the change of inbreeding from year to year or generation to generation. 5, 7, 8, 13, 18, 21, 25, 28

generation interval Average age of dams from a given sub-group of a population (e.g. animals born in a given year). 8, 13, 17, 18, 22, 25

herdbook is a register containing information about a breed on the individuals level. It usually contains the birthdate, the sex and *pedigree* information. Note that the name herdbook is usually used for cattle animals; however, it is often extended to other animals, which is the case in this document. 6

inbreeding (F) For a given individual, its inbreeding coefficient is calculated based on the relatedness of the ancestors and it represents the expected homozygosity of its genome (Wiggans et al., 1995). 5–8, 13, 14, 17, 18, 20–22, 24, 25, 27, 28, 30, 36

introgression defines the process by which gene flow from a different genetic pool enters another genetic pool (Dowling & Secor, 1997). In our case, introgression can be achieved by impregnating a female from a specific breed with sperm from a sire of another breed. 5, 7, 8, 13–15, 20, 22–24, 26, 27, 30

pedigree list of ancestors of an animal. 4–6, 8, 9, 13–15, 17, 28–31, 36, 37

Unix family of open source *OS*. 29

Acronyms

BAS Breed Agriculture Sustainability: An (ad-hoc made) index to evaluate the agriculture sustainability at the breed level, see section 2.1. 7, 10, 28

BFS Swiss Federal Statistique Office. 9, 15, 16

BLW Bundesamt für Landwirtschaft, the Swiss Federal Office for Agriculture. 5, 29

BVO Braunvieh Original: a cattle breed. 20, 24, 25, 27, 28, 30

CSS Cascading Style Sheet is a language used in Web-development and allowing to code the graphical content and physical appearance of a page. 29

DBMS database management system is a software designed to interact between the user and the database. 17, 29

ELECTRE “ELimination and Choice Expressing REality”, an MCDA-technique. 11

EPFL Ecole Polytechnique Fédérale de Lausanne; Swiss Federal Institute for Technology.

5

FAO Food and Agriculture Organization is an agency of the United Nations. 5, 30

FM Franches-Montagnes: a horse breed. 20, 24, 27, 30

GI generation interval. 7, 20

GIS Geographic Information System. 6, 15, 17, 29, 30, 32

HAFL Hochschule für Agrar-, Forst- und Lebensmittelwissenschaften; the School of Agriculture, Forest and Food Science. 5

HTML Hypertext Markup Language is a markup language used for to create Web-pages. 17, 18, 29

LAS Local Agriculture Sustainability: An (ad-hoc made) index to evaluate the agriculture sustainability of a region, see section 2.1. 7, 9, 10, 13, 15, 16, 27

MACBETH “Measuring attractiveness by a categorical based evaluation technique”, an MCDA-technique. 11, 13, 28

MCDA Multi-Criteria Decision Analysis. Families of techniques designed to aggregate criteria. 6, 28

OS Operating System like Windows7 or LINUX-ubuntu. 29, 31

PHP Hypertext Preprocessor is a server-side programming language used in Web development. 17, 18, 29

QGIS An open source *GIS* software. 16, 29

SQL Structured Query Language is a language family used in database to retrieve specific information from data. 18

WFS Web Feature Service is a protocol to create maps as dynamic objects. 18

WMS Web Mapping Service is a protocol to create maps as images. 18

WSL Swiss Federal Institute for Forest, Snow and Landscape Research. 10, 15

XML Extensible Markup Language is a markup language used for to encode documents in the creation of Web-pages. 18

ZIP-code Postal codes. 8, 10, 13–16, 18, 22–24, 26, 27

References

- Alderson, L. (2009). Breeds at risk: Definition and measurement of the factors which determine endangerment. *Livestock science*, *123*(1), 23–27.
- Alderson, L. (2010). Breeds at risk. In *Criteria and classification. report from a seminar* (pp. 16–17).
- Bertaglia, M., Joost, S., & Roosen, J. (2007). Identifying european marginal areas in the context of local sheep and goat breeds conservation: A geographic information system approach. *Agricultural systems*, *94*(3), 657–670.
- BFS. (2014). *Portraits rÃ©gionaux 2014: communes* (Tech. Rep.). Bundesamt fur Statistik. Retrieved from <http://www.media-stat.admin.ch/maps/profile/data/237/en/pdf/Regional-portraits-2014-Communes.pdf>
- BLEMV. (n.d.). Animal genetic resources in germany [Computer software manual]. Bonn, Germany.
- Chee-Sanford, J. C., Aminov, R. I., Krapac, I., Garrigues-Jeanjean, N., & Mackie, R. I. (2001). Occurrence and diversity of tetracycline resistance genes in lagoons and groundwater underlying two swine production facilities. *Applied and Environmental Microbiology*, *67*(4), 1494–1502.
- Costa, E., Bana, C. A., & Vansnick, J.-C. (1994). Macbeth an interactive path towards the construction of cardinal value functions. *International transactions in operational Research*, *1*(4), 489–500.
- DEFRA. (2013). Uk country report on farm animal genetic resources 2012 [Computer software manual].
- Dowling, T. E., & Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annual review of Ecology and Systematics*, *28*(1), 593–619.
- Ellis, W. S., & Starmer, W. T. (1978). Inbreeding as measured by isonymy, pedigrees, and population size in tÃ¶rbel, switzerland. *American journal of human genetics*, *30*(4), 366.
- Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: implications for plant conservation. *Annual review of Ecology and Systematics*, *24*(1), 217–242.
- FAO. (2007a). *Global plan of action for farm animal genetic resources and the interlaken conference* (Tech. Rep.). Author.
- FAO. (2007b). *The state of the world’s animal genetic resources for food and agriculture*. Rome: Commission on genetic resources for food and agriculture, FAO.
- Gandini, G., Ollivier, L., Danell, B., Distl, O., Georgoudis, A., Groeneveld, E., . . . Woolliams, J. (2004). Criteria to assess the degree of endangerment of livestock breeds in europe. *Livestock Production Science*, *91*(1), 173–182.
- Goodchild, M. F. (1996). Geographic information systems and spatial analysis in the social sciences. *Anthropology, Space, and Geographic Information Systems (Spatial Information Series)*, *New York, Oxford*, 241–250.
- Goyache, F., GutiÃ©rrez, J. P., FernÃ¡ndez, I., GÃ³mez, E., Ãlvarez, I., DÃ­ez, J., & Royo, L. (2003). Using pedigree information to monitor genetic variability of endangered populations: the xalda sheep breed of asturias as an example. *Journal of Animal Breeding and Genetics*, *120*(2), 95–105.
- Groeneveld, E., Westhuizen, B., Maiwashe, A., Voordewind, F., Ferraz, J., et al. (2009). Poprep: a generic report for population management. *Genetics and Molecular Re-*

- search*, 8(3), 1158–1178.
- Guitouni, A., & Martel, J.-M. (1998). Tentative guidelines to help choosing an appropriate media method. *European Journal of Operational Research*, 109(2), 501–521.
- Gutiérrez, J. P., Cervantes, I., Molina, A., Valera, M., & Goyache, F. (2008). Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genet. Sel. Evol*, 40, 359–378.
- Houghton, J. (2000). The ipcc report 2001. In *The solar cycle and terrestrial climate, solar and space weather* (Vol. 463, p. 255).
- Joost, S., Colli, L., Baret, P., Garcia, J., Boettcher, P., Tixier-Boichard, M., & Ajmone-Marsan, P. (2010). Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Animal genetics*, 41(s1), 47–63.
- Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5), 230–241.
- Notter, D. R. (1999). The importance of genetic diversity in livestock populations of the future. *Journal of Animal Science*, 77(1), 61–69.
- Price, B., Kienast, F., Seidl, I., Verbrug, P., Ginzler, C., & Bolliger, J. (2013). Spatially explicit modelling of land-use suitability and future land-use pattern for switzerland. In *Abstract volume, 11th geoscience meeting - computational giscience*.
- Roy, B. (1991). The outranking approach and the foundations of electre methods. *Theory and decision*, 31(1), 49–73.
- Swisstopo. (2014). *Swissboundaries3d* (Tech. Rep.). Bundesamt fur Landestopographie. Retrieved from <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/swissBOUNDARIES3D.html>
- Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. *Ecological Economics*, 45(3), 365–376.
- Wiggans, G., VanRaden, P., & Zuurbier, J. (1995). Calculation and use of inbreeding coefficients for genetic evaluation of united states dairy cattle. *Journal of dairy science*, 78(7), 1584–1590.
- Woolliams, J., et al. (1994). Effective sizes of livestock populations to prevent a decline in fitness. *Theoretical and Applied Genetics*, 89(7-8), 1019–1026.
- Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.*, 56, 330–338.

8 Appendices

A Thresholds applied in different countries

A.1 Food and Agriculture Organization (FAO)

The FAO only takes into account the number of individuals. They define the following scale to determine which breeds are at risk.(FAO, 2007b)

CATEGORY	Description
EXTINCT	No longer possible to recreate Becomes absolute when no breeding males/females remain Extinction may be realized well before
CRITICAL	Number of breeding females ≤ 100 ; or Number of breeding males ≤ 5 ; or Population size is ≤ 120 and decreasing and the %age of females bred to males of the same breed $< 80\%$
ENDANGERED	Number of breeding females > 100 and ≤ 1000 ; or Number of breeding males ≤ 20 and > 5 ; or Population size is > 80 and < 1000 and increasing and the %age of females bred to males of the same breed $> 80\%$; or Population size is > 1000 and ≤ 1200 and decreasing and the %age of females bred to males of the same breed $< 80\%$. Further categorization as CRITICAL-MAINTAINED or ENDANGERED-MAINTAINED:critical/endangered populations with active conservation programmes or that are maintained by commercial companies/research institutions.
NOT AT RISK	If none of the above definitions apply and: Total number of breeding females/males $> 1\ 000/20$, respectively; or If Population size $> 1\ 200$ and increasing.

Table 5: Endangerment scale as defined by the FAO

A.2 United Kingdom

Alderson proposes a review of the UK monitoring system (Alderson, 2009). In this system, three parameters are taken into account:

- Numerical: number of breeding females
- Geographical: radius of the circle which entails 75% of the population
- *Inbreeding (F)*: expected average *inbreeding (F)* coefficient in 25 years

Based on these parameters they define the scale shown in table 6

Category	Numerical	Geographical	Inbreeding
Critical	100-300	12.5	30
Endangered	165-500	15	25
Vulnerable	300-900	17.5	20
At Risk	500-1500	20	15
Transitional	1000-3000	25	10

Table 6: Endangerment scale used in the UK

A.3 Germany

Germany bases its monitoring system on the effective population size. They use equation 8 to calculate this parameter but do not mention the way they define their population (i.e. it might be based on method 2 or 3). Its monitoring system is based on the effective population size and the following scale is used:

Category	N_e
Conservation Population	<200
Phenotypic Conservation Population	<50
Monitoring Population	<1000
Non-Endangered Population	>1000

Table 7: Endangerment scale used in Germany

B Software used in *pedigree* analysis

A wide variety of software exist to treat *pedigree* information. We can cite among other a few that specifically computes the above-mentioned parameters.

- PopRep: it is a software developed in Perl using opensource softwares such as gnu-plot, postgresql, latex ... It allows to compute all the above-mentioned parameters. It is integrated into a database system.
- ENDOG: this application is built in C++ and computes almost the same parameters as PopRep
- Coefficient, Inbreeding, Coancestry (CFC): it mainly offers the calculation of inbreeding coefficient as well as a mating optimization in order to minimize the inbreeding coefficient.

C Important parameters used in *pedigree* analysis

Number of breeding animals This simply considers the total number of animals, that may be separated by gender

Inbreeding (F) and its rate of change The inbreeding is a measure of common ancestor of a given individual. It represents the homozygosity expected in an animal genome because of the relatedness of its parents (Ellis & Starmer, 1978) It is calculated according to the following formula:

$$F = \sum [(\frac{1}{2})^{n_1+n_2+1} * (1 + F_A)] \quad (2)$$

(Wright, 1922) where n_1 is the number of generations from the sire to the common ancestor n_2 is the number of generations from the dam to the common ancestor F_A is the inbreeding coefficient of the common ancestor considered.

The rate of change of the inbreeding among generations is defined as follow:

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad (3)$$

where F_t is the average coefficient of a given generation (all animals born in a given year). and F_{t-1} is the average inbreeding coefficient of the previous generation. Groeneveld identifies to ways to define the previous generation (Groeneveld et al., 2009):

- the set of parents of all animals in the generation t
- all animals born one generation earlier based on the average generation interval.

This parameter is often considered as important in this topic in the literature. See for example (Gandini et al., 2004), (Goyache et al., 2003), (Alderson, 2010)

Effective population size (N_e) The effective population size represents the "number of individuals in an ideal population that would have the same genetic response to random processes as a real population of size N" (Ellstrand & Elam, 1993). Groeneveld reports six different methods to determine this parameter (Groeneveld et al., 2009):

- Method 1: $N_e - cens$

$$N_e = \frac{4N_m N_f}{N_m + N_f} \quad (4)$$

where N_m is the number of males and N_f is the number of females. This equation assumes that the mating ratio equals 1. A mating ratio of 1 is in reality almost never achieved since the number of females usually largely exceeds the number of males. The other methods are based on the following equation:

$$N_e = 1/2\Delta F \quad (5)$$

where ΔF is computed using equation 3

- Method 2: $N_e - \Delta F_p$

The ΔF is computed using the parents of all animals of the generation t. It reflects the most recent estimate.

- Method 3: $N_e - \Delta F_g$ this time the generation t-1 is defined as being the animals born one generation interval before generation t. The generation interval can be computed is most commonly defined as the average age of the parents at the birth of their selected offspring. An offspring is considered selected if it has produced at least one progeny.

- Method 4: $N_e - \Delta f_g$

Instead of using ΔF , this time Δf is used. Δf is the rate of change of additive genetic relationship (AGR). AGR is reflects the total average genetic relationship in a group. If inbreeding within herds is high AGR may still be low.

$$\Delta f = \frac{f_t - f_{t-1}}{1 - f_{t-1}} \quad (6)$$

- Method 5: $N_e - ln$

Uses the following equation

$$\Delta F = (-1)bL \quad (7)$$

where L is the generation interval and b is the slope of the regression

- Method 6: $N_e - ecg$

Is based on the following formula

$$N_e = 1/2(1 - (1 - F_t^{1/t})) \quad (8)$$

where F_t is the average inbreeding coefficient of all animals born within the same year and t is the number of years it took to reach F_t which can be calculated using the following formula:

$$t = \sum_{i=1}^n \left(\frac{1}{2}\right)^i \quad (9)$$

where i is the number of generations separating the individual from each known ancestor (n being the total number of known ancestors).

The effective population size is also an often-used parameter to monitor breeds at risk of extinction. See for example (Woolliams et al., 1994), (Gandini et al., 2004), (Goyache et al., 2003), (Alderson, 2010). The question is now to define which method is best suited for the monitoring of a given breed. The answer to this question is not straightforward. Gutiérrez *et al.* argues that method 6 is applicable in most situations (subdivision of population, generation overlap, ...) and that it gives a more stable result than the computation with method 2 and 3, while remaining sensible to recent changes. (Gutiérrez et al., 2008) Groeneveld in turn argues that the answer should not be unique and that the choice of one method should depend on the situation. In the module called PR-Mon CITATION he explains a possible classification tree. The first thing to consider is to see whether there is stratification (i.e. selection within herds). If so, Method 4 should give a larger result and this number should be left out. Then, considering the 4 other methods, and in a will to show the most recent evolution of a breed, the methods with the smallest time window will be chosen. This corresponds to the following sequence of methods: method 5 then 2, 3, 4, 6, 1. By default, method 5 will be chosen, unless N_e is not complete or not stable enough from year to year (more than 50% drop or increase within one year). The sequence is followed until it reaches a parameter that is both complete and stable.

Family size The family size is the number of offspring of an individual that become breeding individuals in the next generation.

Other Goyache et al. also proposes a list of parameters regarding the number of herds. These parameters are not often found in the literature. (Goyache et al., 2003)

D Thresholds and weights used

This section gives a summary of the thresholds and weights that were used in the computation whose results are shown in this paper. They can of course be changed by the user in the future.

For the **pedig-index**:

Criteria	t_1	t_2	weight
Mean inbreeding (last GI)	0.1	0.02	0.7
N_e	50	1000	0.3

Table 8: Parameters for the pedig-index

For the **LAS-index**:

Criteria	t_1	t_2	weight
% change WSL	94	100	0.2
Demographical balance	0	3	0.1
% social assistance	5	2	0.2
% farmer	1	16	0.1
% grazing surface	6	30	0.05
Evolution number farmers	0	10	0.1
% less 19	3	10	0.2
% more 65	20	4	0.05

Table 9: Parameters for the LAS-index

For the **global-index**:

Criteria	t_1	t_2	weight
pedig-index	0.15	0.8	0.5
introg-index	0.1	0.02	0.2
geog-index	12	50	0.1
LAS-index	0.4	0.8	0.2

Table 10: Parameters for the global-index

E Details of the land use scenario of the WSL

Each scenario models different amount and spatial allocation of 6 landuse types:

- Closed Forest
- Open Forest
- Overgrown (abandoned) areas
- Agriculture - Arable
- Agriculture - Meadows
- Urban/sealed areas

The six available scenarios are

Trend

- Linear interpolation of 1985-1997-2009 trend in growth (or reduction) of all land use classes

A1

- Urban demand 315754 ha = BfS "Tiefe" population scenario, mean per capita urban area demand
- Restrictions on where urban growth can occur through the existing "Bauzone"

A2

- Urban demand 502118 ha (= BfS "Hohe" population scenario and 95% Upper CI limit for urban area demand per capita)
- No further spatial restrictions
- Weighting of urban suitability to reflect regionalisation and improved public transport connectivity in more rural areas = areas currently villages/small towns with public transport accessibility but at a low ranking (regionalisation scenario)

B1

- Urban demand -1400 ha from 2009 level (= "Stagnation" scenario from Fahrländer Partner)
- Restrictions on where urban growth can occur through the existing "Bauzone"
- Restrictions preventing any conversion from Agriculture-Meadows in areas with elevation over 900m asl
- Increased demand for agriculture

B2

- Urban demand 359700 ha (= BFS "Mittelere" scenario and mean urban surface area per capita)
- No "Bauzone" restrictions on urban growth
- Restrictions preventing any conversion from Agriculture-Meadows in areas with elevation over 900m asl
- Weighting of urban suitability to reflect regionalisation and improved public transport connectivity in more rural areas = areas currently villages/small towns with public transport accessibility but at a low ranking (regionalisation scenario)
- Increased demand for agriculture

F Install GenMon-CH on a new server

Here is the description for Ubuntu

Install Apache and PHP, then install postgresql with its spatial extension postgis, and link php with postgresql (usually done by uncommenting the line “extension=php_ pgsqldll” in the php.ini file)

Put the genmon-ch.zip somewhere in the file. In our case, the directory was /home/lasigadmin/genmon-ch. If you change this directory, you have to change it in the codes when it appears.

Define the path to this directory using an alias in the server configuration file.

Restore the database by running the .backup in a sql query (the backup is of type plain using insert into command and can thus be run as a sql file). Depending on the name, user and password you might want to change the connection to the database in the connectDataBase.php.

Download Geoserver. Publish the layer plzo_ plz from the PostGIS (first create a new store and define a connection to the postgresql database then add a new layer to this store. The SRC is EPSG:3857 and you can calculate the footprint based on data. Save this layer as plzo_ plz.

Regarding PopRep, the code is already in the .zip file. Nevertheless, you need to download a few things...

First, the perl module:

```
[CORE] Cache::Memcached Compress::Zlib Config::IniFiles Config::Auto CPAN Data::Dumper
Date::Calc Digest::MD5 File::Basename List::Util List::MoreUtils Scalar::Util Storable
Sys::Syslog Text::ParseWords XML::Parser XML::Writer XML::DTDParse
```

```
[TK] Tk 804.024 Tk::BrowseEntry Tk::DateEntry Tk::MListbox Tk::ObjEditor Tk::ObjScanner
Tk::MatchEntry Tk::Table
```

```
[DATABASE] DBD::CSV DBD::Pg 1.30 DBI 1.35 DBIx::Recordset SQL::Statement Text::CSV_
XS
```

```
[I18N] Encode Locale::Maketext Locale::Maketext::Lexicon
```

```
[NETWORKING] Net::EasyTCP 0.26 Net::FTP
```

```
[MISC] Statistics::Descriptive Template::Plugin::DBI Term::ReadKey Tie::IxHash
Time::HiRes Spreadsheet::WriteExcel Spreadsheet::ParseExcel Unicode::String Unicode::Map
File::MMagic
```

```
[WEB] HTML::Template Apache::Session
```

```
[ENCRYPTION] Crypt::CBC Crypt::Blowfish
```

```
[COMPRESION] Compress::Zlib
```

```
+ forkmanager
```

Then install latex and texlive-full (full will prevent you from downloading all packages and fonts separately which is tiring). Make sure the default paper size is A4 in the texconfig (sudo texconfig).

Finally you need to define a permanent environment variable for the path to the apiis file. You also need to define this in the envvar file of the apache configuration files.