Quantitative approaches to the analysis of sedimentary DNA to understand past biodiversity and ecosystem functioning

Wentao Chen

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biodiversité-Ecologie-Environnement

Arrêté ministériel : 25 mai 2016

Présentée par

WENTAO CHEN

Thèse dirigée par **Pierre TABERLET**, CHERCHEUR CNRS, CNRS
et codirigée par **Fabien ARNAUD**, Directeur de Laboratoire, USMB

préparée au sein du **Laboratoire Laboratoire d'ECologie Alpine** dans **l'École Doctorale Chimie et Sciences du Vivant**

Approches quantitatives de l'analyse de l'ADN sédimentaire pour comprendre la biodiversité et le fonctionnement des écosystèmes dans le passé

Quantitative approaches to the analysis of sedimentary DNA to understand past biodiversity and ecosystem functioning

Thèse soutenue publiquement le **11 février 2019**,
devant le jury composé de :

Monsieur PIERRE TABERLET
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES, Directeur de thèse
Madame LAURA PARDUCCI
PROFESSEUR ASSISTANT, UNIVERSITE D'UPPSALA - SUEDE, Rapporteur
Madame INGER GREVE ALSOS
PROFESSEUR, UNIVERSITE DE TROMSØ - NORVEGE, Rapporteur
Monsieur PHILIPPE CHOLER
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES, Examinateur
Monsieur DIDIER GALOP
DIRECTEUR DE RECHERCHE, CNRS DELEGATION OCCITANIE OUEST, Examinateur
Monsieur RACHID CHEDDADI
DIRECTEUR DE RECHERCHE, CNRS DELEGATION LANGUEDOC-ROUSSILLON, Président

Communauté UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biodiversité-Ecologie-Environnement

Arrêté ministériel : 25 mai 2016

Présentée par

WENTAO CHEN

Thèse dirigée par **Pierre TABERLET (CSV)**, CHERCHEUR CNRS, CNRS
et codirigée par **Fabien ARNAUD**, Directeur de Laboratoire, USMB
et coencadrée par **Gentile Francesco Ficetola**, Professor, Université de Milan
préparée au sein du **Laboratoire Laboratoire d'ECologie Alpine** dans **l'École Doctorale Chimie et Sciences du Vivant**

Approches quantitatives de l'analyse de l'ADN sédimentaire pour comprendre la biodiversité et le fonctionnement des écosystèmes dans le passé

Quantitative approaches to the analysis of sedimentary DNA to understand past biodiversity and ecosystem functioning

Thèse soutenue publiquement le **11 février 2019**,
devant le jury composé de :

Monsieur PIERRE TABERLET (CSV)
Directeur de Recherche, Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble Alpes, Directeur de thèse
Madame LAURA PARDUCCI
Professeure assistante, Department of Ecology and Genetics, Plant Ecology and Evolution, Uppsala university, Rapporteur
Madame INGER GREVE ALSOS
Professeure, Department of Natural Sciences, the Arctic University of Norway, Rapporteur
Monsieur PHILIPPE CHOLER
Directeur de Recherche, Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble Alpes, Examinateur
Monsieur DIDIER GALOP
Directeur de Recherche, Laboratoire GEODE, Université Toulouse-Jean-Jaurès, Examinateur
Monsieur RACHID CHEDDADI
Directeur de Recherche, Institut des Sciences de l'Evolution-Montpellier, Université de Montpellier, Examinateur

Quantitative approaches to the analysis of sedimentary DNA to understand past biodiversity and ecosystem functioning

PhD candidate: Wentao Chen

Supervised by
Dr. Pierre Taberlet

Co-supervisors:
Dr. Fabien Arnaud
and
Dr. Gentile Francesco Ficetola

December 2018

To my parents

Πολλὰ τὰ δεινὰ κοὐδὲν ἀνθρώπου δεινότερον πέλει.

Many things cause terror and wonder, yet nothing

is more terrifying and wonderful than man.

Sophocles, *Antigone*

Acknowledgements

I would like to thank the support of many people: Gentile Francesco Ficetola, Pierre Taberlet and Fabien Arnaud for their supervising and for the many fruitful discussions; Pierre Sabatier, Manon Bajard, Erwan Messager, Jérôme Poulenard, and Philippe Choler for their inspiring discussions and comments; Rachid Cheddadi for evaluating my work and giving precious suggestions through the course; I am grateful to Ludovic Gielly and Delphine Rioux for their help during laboratory work; Frédéric Boyer and Céline Mercier for their help with bioinformatics; Olivier Lontin for general informatics support; Jesùs Maravez for numerous tea-time discussions; François Pompanon and Christian Miquel for their help with life in the LECA; Florence Sagnimorte and Agnès Agarla for logistic support; I acknowledge the many people in the LECA with whom I have passed so much joyful office and leisure time through my years in the lab: João Braga, Ceres Barros, Thibaut Capblancq, Marta de Barba, Roberto Geremia, and all the other people working in the LECA. I am thankful to the emotional and intellectual support of many friends, most importantly, Mao Zhun, Zhu Huaxiang, Zhang Linran, Gilles Nicolet, Liu Li. Special thanks must be given to my girlfriend, Li Hanyang, who has been accompanying and supporting me for the last two years, and without whom my PhD journey would have been much less enjoyable.

Contents

Chapter 1. Introduction

1.1 Biodiversity and ecosystem functioning in the Anthropocene

Human activities have been exerting significant impacts on the environment since the dawn of the Neolithic, with the advent of agriculture and related biomass burning (Glikson & Groves, 2016; Steffen, Grinevald, Crutzen, & Mcneill, 2011). These activities have become a major driving force of Earth system change, even rivaling the forces of nature, especially since the Great Acceleration (mid-20th to early-21st century) (Steffen, Crutzen, & McNeill, 2007; Waters et al., 2016). This observation has led to the concept of Anthropocene (Crutzen, 2002), which describes a geological epoch of human-driven global changes. These changes include but are not limited to alternations in land use and land cover, urbanization, globalization, atmospheric composition, riverine flow, nitrogen cycle, carbon cycle, physical climate, and biodiversity (Steffen et al., 2007).

Biodiversity and ecosystem functioning changes under anthropogenic forcing are one of the critical aspects of global change, for they threaten the sustainability of our own species and others' (Ruddiman, 2013; Steffen, Persson, et al., 2011; Vitousek, Mooney, Lubchenco, & Melillo, 1997). Human impacts on biodiversity and ecosystem functioning can be perceived at different temporal and spatial scales (Brose, Hillebrand, & Brose, 2016; Loreau et al., 2001). Nevertheless, in a global scale, species extinction rate has increased by as much as 1,000 times over background rate (Reid et al., 2005). The global Live Planet Index, an index that measures the rate of change in global biodiversity by aggregating population data among vertebrate species (Collen et al., 2009), shows a decline of around 60% between 1970-2012 (WWF, 2016). Overall, current ecosystem trends are threatening the well-being of a large part of people of their basic materials for a good life health, security, etc., especially for those in poor populations (Marc, Babu, & Hamilton, 2005).

Human activities affect biodiversity via multiple direct or indirect drivers (Vitousek et al., 1997). Five major Anthropogenic drivers were identified: land use change, climate change,

invasive species, over-exploitation and pollution (Millenium Ecosystem Assessment, 2010). Their direct and indirect effects are summarized in 1-1.

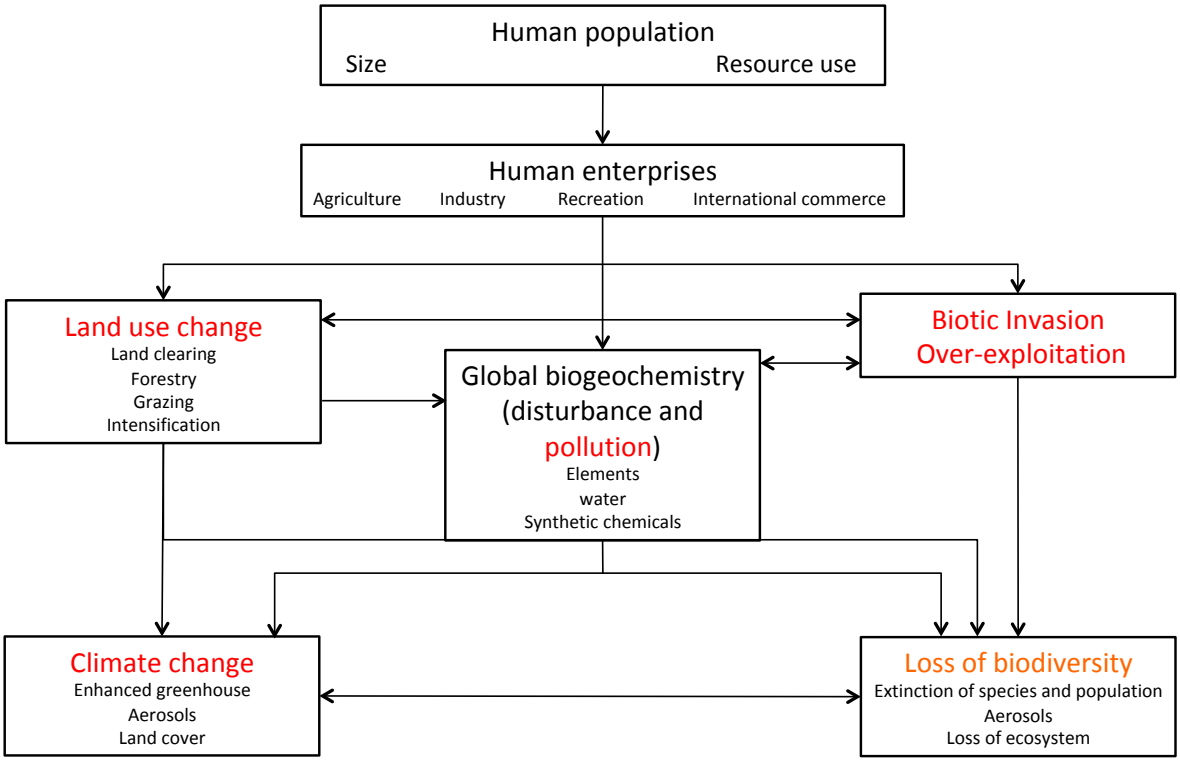


Figure 1- 1 A conceptual model illustrating the direct and indirect effects of human activities on ecosystems, adjusted from *(Vitousek et al., 1997)* according to *(Millenium Ecosystem Assessment, 2010)*

The impacts of different drivers and their relative importance vary depending on the biome. For instance, on mountains, land use change has been a continuing and highest impact on biodiversity over the past 50-100 years, whereas in polar environment, it was climate change that has been the largest and increasingly important driver (Millenium Ecosystem Assessment, 2010).

Given the pervasive and complex impacts of human activities, understanding and forecasting the impacts of human activities on biodiversity and ecosystem functioning is not only an interesting intellectual challenge to ecologists, but also a pressing question

crucial to the sustainability of human well-being (Loreau et al., 2001; McGill, Dornelas, Gotelli, & Magurran, 2015). Retrospective studies on biodiversity and ecosystem functioning (usually under the term 'paleoecology') can provide indispensible information upon how ecosystems react to environmental changes. With such studies, one can establish the baseline and spatio-temporal variability of an ecosystem with supposingly less disturbed conditions, formulate cause-effect explanations and test hypotheses through inference from multivariate data, and even seek complex system behaviors from such datasets (Dearing, Battarbee, Dikau, Larocque, & Oldfield, 2006). Consequently, knowledge gained from retrospective studies will help forecast ecosystem changes that may be vital to human well-being.

1.2 eDNA metabarcoding for the study of past biodiversity

Environmental DNA (eDNA) is an emerging tool for biodiversity study. It can be defined as a complex mixture of DNA from various organisms contained in environmental samples (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012), such as soil (Taberlet, Prud'Homme, et al., 2012; Yoccoz et al., 2012), water (Dejean et al., 2012; Muha, Rodríguez-Rey, Rolla, & Tricarico, 2017), feces (Pompanon et al., 2012), sediments (Giguet-Covex et al., 2014; Parducci, Jørgensen, et al., 2012; Willerslev et al., 2014), and even some unconventional sources like spider webs (C. C. Y. Xu, Yen, Bowman, & Turner, 2015). In the study of past ecosystem, eDNA can be collected from lake sediment (Giguet-Covex et al., 2014), permafrost sediment (Jørgensen et al., 2012), and archeological remains like midden (Seersholm et al., 2016). From eDNA, it is possible to retrieve taxonomic, genetic and functional information of ecosystem by analyzing the DNA sequences. Instead of targeting single species, such studies often aim to obtain information about multiple taxa (e.g. all the vascular plants). This can be achieved by using targeted PCR or shotgun sequencing, and such approach is called "DNA metabarcoding" [(Taberlet, Bonin, Zinger, & Coissac, 2018), see **1.3** for details]. Here we focus on the use of sedimentary DNA metabarcoding to assess past (mainly vascular plant) biodiversity and ecosystem functioning, as sediments are the most wildly collected source of eDNA from

the past, and provide stratified information that allow the reconstruction of long-term ecosystem dynamics.

Traditionally, stratified information of past biodiversity can be inferred from pollen and macrofossils found in sediments. Previous studies comparing data from pollen, macrofossils and eDNA have identified some of the advantages and disadvantages of eDNA over the two traditional sources (Alsos, Sjörgren, et al., 2016; Jørgensen et al., 2012; Parducci et al., 2013, 2015; Pedersen et al., 2013). Each data source has its unique taxa undetected from other sources. However, the overlap between eDNA and macrofossils is higher than that between eDNA and pollen. eDNA has also origins similar to macrofossils, and therefore provide localized biodiversity information (Alsos, Lammers, Yoccoz, & Edwards, 2018). In contrast, pollen data is biased towards wind-dispersed pollen susceptible to long-distance transportation, resulting in a regional view of plant communities. Taxonomic resolution is often higher in eDNA and macrofossils than in pollen. To obtain higher taxonomic resolution, markers developed specifically for a given taxonomic group can be applied to eDNA [e.g. the Poac01 marker for the grass family Poaceae (Taberlet et al., 2018)], which is not possible for pollen and macrofossils. Furthermore, taxonomic assignment of eDNA sequences can be (and is nearly always) automatized, unlike those of pollen and macrofossils, which require expertise in the morphology of the target group and a large amount of human interventions and workload.

One of the difficulties of sedimentary DNA-based study of past biodiversity was related to the unknown taphonomy of DNA in sediment. Recent studies revealed that most sedimentary DNA reflected plant communities very close to the sampling site. For example, Alsos, Lammers, Yoccoz, Jørgensen, et al., 2018 found that up to 73% of sedimentary plant DNA sequences matched the taxa recorded in vegetation surveys within 2m from the lake shore, suggesting the highly localized signals of eDNA, similar to the taphonomy of macrofossils. However, some taxa detected by eDNA were not recorded in the vegetation surveys but common in the region, suggesting DNA transported from up to a

few kilometers away could possibly be archived in lake sediments. Furthermore, DNA can also be transported by a river that gathers sediment from its drainages.

Imperfect detection in eDNA is another issue worth mentioning. Like most detection methods, imperfect detection is an unavoidable feature of eDNA metabarcoding, which means that both false negatives (FNs) and false positives (FPs) occur (Ficetola et al., 2015). To limit FNs, replicated analyses are performed to ensure the detection of taxa with low detection rate. On the other hand, controlling for FPs is less straightforward, since multiple approaches are needed (Ficetola, Taberlet, & Coissac, 2016; Taberlet et al., 2018). First, multiple laboratory practices are needed to prevent contaminations (e.g. dedicated laboratory rooms for DNA extraction and amplification, discarding the outer parts of sediment samples and extraction blanks). Second, negative controls are performed to detect potential contamination in the extraction and amplification steps. Third, sequences are filtered in the data analysis and samples are compared with controls, removing any suspicious sequences. Finally, site occupancy-detection models (Mackenzie et al., 2002; Schmidt, Kéry, Ursenbacher, Hyman, & Collins, 2013) can be applied to the sequence dataset to estimate the rates of FP and FN, and also the probabilities of presence for each taxon in each sample (see **Chapter 2**). In sedimentary DNA metabarcoding using the PCR-based approach (**1.3**), imperfect detection is a particularly crucial issue. In fact, the ancient DNA found in sediments is often highly degraded and remains in low amount. Therefore large numbers of PCR cycles and of replicates are needed to increase detection rates. These practices can potentially lead to increased FP rate, by introducing and amplifying contamination and sequencing errors. Consequently, all the practices named above are usually adopted to ensure the reliability of DNA analysis results, with eventually additional steps to further avoid spurious results (Ficetola et al., 2015, 2016). In shotgun-sequencing- and capture-based approaches of ancient DNA metabarcoding, an additional verification of the ancient origin of DNA can be performed exclude the impact of modern DNA contamination, by looking at damage on both ends of the fragments, as ancient DNA molecules have a higher proportion of damage near their extremes (Hofreiter, Jaenicke, Serre, Haeseler, & Pääbo, 2001; Mikkel W. Pedersen et al., 2016; Weiß, Dannemann, Prüfer, & Burbano, 2015).

1.3 Workflow in eDNA metabarcoding for past biodiversity study

Figure 1.2 describes the main steps of typical eDNA metabarcoding for past biodiversity study.

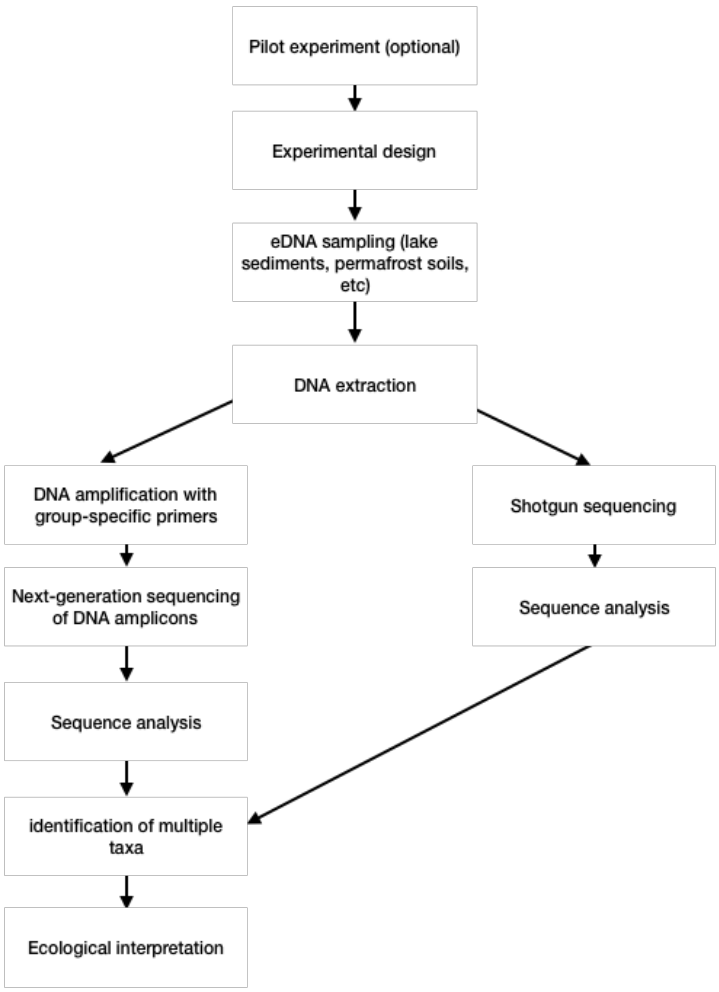


Figure 1.2 Workflow of environmental DNA metabarcoding for past biodiversity study, simplified from (Taberlet et al., 2018)

In eDNA metabarcoding for past biodiversity study, samples are usually collected from natural records such as lake sediments. These raw materials are often chronologically arranged in nature, with various temporal resolutions, which can be up to annual level for varved materials (Mills et al., 2016). However, for most eDNA studies, a relatively large amount of raw material must be used to collect enough DNA contents [e.g. ca. 14 g for lake sediment in (Giguet-Covex et al., 2014) and 6 g for permafrost sediment in (Sønstebø et al., 2010)]. The time scale covered by a sediment sample of a given weight is not constant

through the whole core, due to varying sedimentation rate. An optimal temporal resolution therefore cannot always be reached, and can vary from decadal to centennial. The varying temporal resolution may lead to some issues in the ecological analysis and interpretation step, which we will discuss further in par. **1.4**.

As mentioned in par. **1.2**, there are two strategies to identify multiple taxa in a study. The first one involves the use of PCR-based amplification of one or multiple DNA metabarcodes (i.e. short, standardized group-specific genetic markers for the taxonomic identification of multiple taxa). Metabarcodes for many major taxonomic groups [e.g. Sper01 for seed plants, Mam02 for mammals, (Taberlet et al., 2018)] are already available. New metabarcodes can be designed for specific purposes via dedicated computer software, such as ecoPrimers (Riaz et al., 2011). Because the PCR step is prone to magnify contaminations and to introduce new errors, it must be meticulously performed to limit these errors, especially including negative PCR controls, which will also serve as a starting point to estimate the FP rate in the SODMs (see par. **1.2** and **Chapter 2**).

In the sequence analysis step, vigorous pipeline must be built to properly identify, merge, and filter the sequence reads resulting from next-generation sequencing. Software dedicated to these purposes is also available, such as the *obitools* package (Boyer et al., 2016).

In order to identify all taxa in the targeted group of organisms with a highest-possible taxonomic resolution, an eDNA-based past biodiversity study needs a comprehensive reference genetic database including reference sequences of all the taxa that could be potentially present in the study area (Taberlet, Coissac, et al., 2012). Previously, such studies usually relied on (incomplete) global databases reduced to locally present taxa (e.g. Pansu et al., 2015) or incomplete local databases (e.g. Alsos, Lammers, Yoccoz, Jørgensen, et al., 2018), which might have limited the taxonomic resolution and the comprehensiveness of DNA metabarcoding. Fortunately, some comprehensive genome databases will be soon available [e.g. the Phyloalps project collecting the chloroplast genomes of all vascular plants present in the European Alps (Coissac, Hollingsworth,

Lavergne, & Taberlet, 2016)]. For PCR-based metabarcoding, the quality of the metabarcode is also a limiting factor to the taxonomic resolution. Usually, all the taxa in the targeted group cannot be amplified by a metabarcode, and some of the amplified taxa cannot be well-identified, which means some metabarcode sequences cannot discriminate among some taxa due to intra-taxon polymorphism (Ficetola et al., 2010). Taberlet et al., 2018 described the currently available metabarcodes for various taxonmic groups, ranging from bacteria (Cyanobacteria), Archaea, eukaryote, to more specific groups such as seed plants, fungi, and some specific plant families, e.g. Asteraceae (the daisy family). Species-level resolution is not always available even for the most specific metabarcode. For example, the Poac01 metabarcode for the Poaceae (the grass family), originally introduced by (Baamrane et al., 2012), can only identify 52.8% of the species to the species level. Long metabarcodes potentially allow a better taxonomic resolution, but ancient DNA found in sediments is often highly degraded. In such case, longer metabarcodes are not suitable. The Sper01 metabarcode (corresponding to the g/h primers in Taberlet et al., 2007), with a minimum length of 10 bp and a maximum of 220 bp, is widely used for analyzing seed plants in degraded templates, and also used in **Chapters 3** and **4** to identify terrestrial plants. It has a coverage of 98.8% for the target group, and can identify 21.5% of the species to the species level, 36.9% to the genus level, and 77.4% to the family level. Given a genome database and a metabarcode, a reference database specific for the chosen metabarcode can be generated, using in silico PCR programs, such as ecopcr (Ficetola et al., 2010), that simulate PCR amplification on a set of DNA sequences. Finally, taxonomic identification of sequences can be done by alignment using general-purpose programs like blastn (Chen, Ye, Zhang, & Xu, 2015), or programs specifically designed for eDNA metabarcoding like ecotag (Boyer et al., 2016).

The second strategy for eDNA metabarcoding is direct shotgun sequencing on the DNA extraction without the step of PCR amplification on a metabarcode. This strategy avoids the FP issues introduced by the PCR step, namely the magnification of contaminant, and PCR errors such as chimeras (Taberlet et al., 2018). It also allows simultaneously targeting all types of organisms, from bacteria to mammals. It offers a straightforward way to roughly estimate the relative abundances of different types of organism by counting the

corresponding reads, which is rarely possible for the PCR-based strategy, due to biased amplification (Taberlet et al., 2018). As mentioned above, the possibility to authenticate the ancient origin of eDNA in the shotgun-based strategy also provides a great advantage over the PCR-based one. However, the shotgun sequencing approach is more difficult to perform, and more expensive than the PCR-based one. Besides, if the target group has a relatively low biomass among others, very few taxonomically informative reads originated from such group can be retrieved from shotgun sequencing. For example, in Pedersen et al., 2016, out of more than one billion reads generated from 32 sedimentary DNA samples, only 2,596 reads were assigned to metazoans at the genus or family levels. Therefore, a reasonable sequencing effort may not be able to provide enough information of the target groups if they are not abundant or are underrepresented in the DNA template. Bioinformatic analysis for shotgun sequencing results is easier than for the PCR-based approach, thanks to the advantages described above. Currently, the "Holi" computational pipeline developed in Pedersen et al., 2016 is available for such analysis, which gives the reads equal chance to align to a database containing the vast majority of organismal sequences.

Another possibility provided by the shotgun-sequencing strategy is to infer functional diversity using the genomic information directly extracted from the sequence fragments--the so called metagenomics approach, mostly applied to the ecology of microorganisms (Handelsman, 2004; J. Xu, 2006). This approach is not discussed in the present thesis, as it seldom targets macroorganisms like vascular plants, and requires a largely different set of data analysis tools.

1.4 Ecological analysis for sedimentary DNA metabarcoding data – quantitative approaches

Sedimentary DNA metabarcoding is potentially an excellent tool for ecological studies, for it provides stratigraphic information of localized past biodiversity with high taxonomic resolution (**1.2**). As paleoecological records, sedimentary DNA offers invaluable

opportunities to test ecological hypotheses in large temporal span and in multiple levels, ranging from population genetics (Parducci et al., 2012) to community ecology (Capo et al., 2016), and to ecosystem functioning (Giguet-Covex et al., 2014). Therefore the last decade sees a growing number of sedimentary-DNA-based studies on past biodiversity, with increasingly complex molecular bioinformatics approaches for data production and filtering. On the other hand, despite the rich and ever-growing arsenal of numerical/quantitative techniques designed for deal with increasingly complex data analysis scenarios, very few of these techniques were adopted for analyzing sedimentary DNA data, apart from some applications of unconstrained ordination (e.g. principle component analysis and nonmetric multidimensional scaling) to extract main trends of a dataset (Pansu, De Danieli, et al., 2015; Pansu, Winkworth, et al., 2015; Zimmermann et al., 2016), and their constrained counterparts (mostly redundancy analysis) to assess the effects of environmental factors (Pansu, Giguet-Covex, et al., 2015). Failing to adopt appropriate tools from such arsenal may prevent harnessing the full potential of sedimentary DNA, or worse, lead to erroneous conclusions. For example, it is widely acknowledged that autocorrelation is a common feature of most spatial/temporal ecological data, and modeling of autocorrelation is required to correctly estimate e.g. environmental drivers behind ecological patterns (Dormann, 2007; Ives & Zhu, 2011). To date, the issue of spatial/temporal autocorrelation is rarely considered in the eDNA literature apart from a few cases (Dougherty et al., 2016; Ficetola, Poulenard, et al., 2018).

The present thesis aims to demonstrate how existent numerical/statistical approaches can improve the analysis of sedimentary DNA data and provide key ecological insights that cannot be obtained otherwise. This section is dedicated to introduce some of these methods that are most relevant to sedimentary DNA data analysis.

1.4.1 Defining the operational units of eDNA data for ecological analysis

Ecological analysis on eDNA data is about manipulating the presence/absence and/or abundance information of taxa, usually estimated from numbers of reads or of positive replicates. Upon tackling the analysis, a crucial question comes into mind: what counts for a “taxon” in eDNA data? For data obtained by the PCR-based strategy, the answer is sometimes straightforward: any unique sequence assigned to a certain taxonomic level

counts as a "taxon" or a molecular operational taxonomic unit (MOTU, *sensu* Pansu et al., 2015). There are cases where several different yet very similar sequences behave (nearly) identically in the dataset (i.e. having the same distribution among samples), suggesting they could represent amplification errors or intraspecific variation, thus they can be safely combined into a MOTU. Alternatively, MOTUs can be defined by clustering the sequences based on distance/similarity, an approach usually adopted in microbiological and mycological studies (Ryberg, 2015; Schloss & Handelsman, 2005).

For data from the shotgun sequencing strategy, however, defining MOTUs becomes a bit trickier, as even sequence fragments with nothing in common can potentially be originated from the same organism, and the taxonomic information in sequence fragments can vary across different locations on the genome. These features of shotgun-sequencing metabarcoding data can certainly mean more useful information, but can also make further statistical analyses much harder, especially for those frequently used to infer community-level variability (see *1.4.3*). Fortunately, many ecological problems can be tackled without defining such units. For instance, Pedersen et al., (2016) used sedimentary DNA metabarcoding with a shotgun-sequencing approach to investigate the viability in the ice-free corridor in North America for human colonization. In this case, merely the presence/absence of some key taxa (e.g. bison) in a given period can provide crucial evidence to test the hypotheses in question. A potential solution to the unit-defining problem in the shotgun sequencing strategy is to cluster the sequences by their taxonomic identities to a fix level. For example, to describe the development of conifer forest in a mountain area, one can combine all sequences assigned to the genus *Pinus* (pine trees) or lower (e.g. *Pinus cembra*, the Swiss stone pine) into a cluster *Pinus* and treat this cluster as an operational unit, and similarly for other conifer taxa. This solution would sacrifice a large amount of information gained by a shotgun sequencing approach. Future studies may develop new statistical methods to fully take advantages from the rich information provided by shotgun-sequencing metabarcoding.

1.4.2 Site occupancy-detection models

For most research projects, one can be contented with a list of specific taxa and their presence/absence in each sample to start ecological analyses. In some projects, the main purpose of looking at the eDNA is to tell if certain taxa were present in some particular periods (e.g. Giguet-Covex et al., 2014; Pedersen et al., 2016). One needs to estimate the reliability of such presence/absence in this case, as the detections are often imperfect (1.2). Assessments for this purpose are usually done in a probabilistic manner. The most widely used probabilistic assessment approach is through site occupancy-detection models (SODMs, or site occupancy models, SOMs).

SODMs involve the modeling of one or several taxa's state of presence/absence (site occupancy) in each sample (or site) as a probability φ. When present, a taxon is detected with probability p_{11} (i.e. the TP rate) and not detected with false negative (FN) probability $1\text{-}p_{11}$. Eventually, when false positives are considered (often a necessity in eDNA-based metabarcoding, see 1.2), an absent taxon can be falsely detected with probability p_{10} (i.e. the FP rate). By fitting a statistical model that links these parameters to observed data (by a maximum-likelihood- or a Bayesian approach), one can estimate those parameters, as well as the probability of the actual presence/absence of each taxon in each sample (Ficetola et al., 2015; Kéry, 2010; Lahoz-Monfort, Guillera-Arroita, & Tingley, 2016; Schmidt et al., 2013). FN/FP errors can occur at both the sampling and the PCR amplification stages, extending SODMs with a multi-level FN+FP setting (Guillera-Arroita, Lahoz-Monfort, van Rooyen, Weeks, & Tingley, 2017). However, as the authors demonstrated, errors rates at the two stages are not identifiable without information other than the survey prone to FN/FP itself, therefore these multi-level SODMs require information from a survey method without ambiguous detections, and from a calibration experiment. Unfortunately, in sedimentary DNA surveys, such information is rarely available.

The results of SODMs provide key ecological information. For instance, the site occupancy probability φ is a parameter of ecological significance, which is used, to name a few, to trace the shift in a species' presence through time for assessing changes in its status

(Schmidt et al., 2013), and to estimate the habitat patchiness of a species (Bailey, Simons, & Pollock, 2004). On the other hand, the posterior probability of presence of each sample, a value that can be calculated from model parameters and observed data, may serve as a measure of target DNA abundance (Ficetola et al., 2015, 2016; Furlan, Gleeson, Hardy, & Duncan, 2016; Lahoz-Monfort et al., 2016).

Covariates can be integrated into SODMs in order to take into account environmental factors that may affect the presence of species and/or the preservation of eDNA (e.g. Dougherty et al., 2016; Willoughby, Wijayawardena, Sundaram, Swihart, & DeWoody, 2016). Furthermore, additional random processes can be specified in SODMs to model ecological mechanisms. For instance, Olajos et al., 2017 used SODM to estimate the colonization date of whitefish in two Scandinavian lakes, based on lake sediment DNA. They modeled the colonization/extinction of whitefish in each lake as random processes that eventually affected the presence/absence of the species, and presented their results as posterior probability distributions of whitefish DNA presence.

In sedimentary DNA metabarcoding, as in all paleoecological records extracted from sediment cores, samples are chronologically arranged so that data corresponding to adjacent samples may be autocorrelated (temporal autocorrelation). In eDNA biodiversity surveys, as in traditional surveys, samples are taken from multiple spatially interconnected sites, thus also the autocorrelation among data from nearby sites (spatial autocorrelation). Such autocorrelations need to be treated properly in order to obtain unbiased estimation of site occupancy, TP rate, FP rate, etc. **Chapter 2** introduces a SODM setting (conditionally autoregressive model, or CAR model) adapted for eDNA metabarcoding studies that may take into account both temporal and spatial autocorrelations in a unified manner.

1.4.3 Measuring abundance

For eDNA-based researches targeting a single species or a few number of species, a quantitative PCR approach (real-time recording of accumulating DNA sequences over the course of the PCR amplification process, (Higuchi, Fockler, Dollinger, & Watson, 1993)) is usually enough to quantify the DNA abundance of the species in question (e.g. Dougherty et al., 2016; Dunker et al., 2016). In the case of traditional PCR that is the most popular approach for multiple-taxa metabarcoding, real-time quantification is not an option. An alternative approach is to use the number of final sequence reads as an index of abundance. However, doubts have been casted to such approach, as biases such as differences in the shedding/transferring of DNA among taxa, gene copy number variation (Vasselon et al., 2018) and primer biases (Elbrecht & Leese, 2015; Nichols et al., 2018) can render cross-taxon comparison invalid, or at least demand meticulous corrections. Nevertheless, comparisons among different samples for the same taxon would be possible, given that the stochastic nature of DNA sequence fragment numbers, amplified by the PCR process, is taken into account and controlled by adequate data transformation. Some authors used fourth-root transformation on the relative proportion of eDNA reads (Zimmermann et al., 2016), which could be an effective way to quantify differences in relative abundance of a single taxon among samples. Others took advantages from a multiple-replication experiment setting, taking the number of positive replications as a measure of relative abundance, in addition to the number of reads (Pansu, Giguet-Covex, et al., 2015). This approach is based on the assumption that the probability of a sequence being amplified in a given PCR replication is positively correlated with its abundance (Furlan et al., 2016). In all, given the many uncertainties in the meaning of sequence abundance (Taberlet et al., 2018), any abundance measurement based on eDNA metabarcoding data should be interpreted with cautious.

1.4.4 Quantifying species diversity

Species diversity is generally considered as a major determinant of ecosystem stability and functioning (Loreau et al., 2001; Loreau & de Mazancourt, 2013). Quantifying species

diversity based on eDNA is therefore a highly rewarding challenge. As discussed in *1.4.3*, uncertainties remain in the meaning of sequence abundance and in the representativeness of number of reads. Several data analysis approaches can be adopted to mitigate the unfavorable effects of those uncertainties. As in traditional surveys and pollen-based paleoecological studies, rarefaction analysis (generating a curve relating the number of species to the number of samples through random resampling) can be performed to assess species richness from limited samples (e.g. Bellemain et al., 2013; Zimmermann et al., 2016). Recent advances in statistics also provide opportunities to cope with uncertainties pertained to eDNA metabarcoding data. For example, the Hill number is a parameterized diversity index with increasing popularity in ecology (Chao, Chiu, & Jost, 2014). It generalizes the three most important diversity indices: species richness (i.e. number of species), the Shannon index, and the Simpson index, by varying the value of a parameter q, which specifies the importance attributed to rare species. This would be very useful to limit the impact of artifact MOTUs. It is even possible to calculate the Hill number from a species (taxonomic, phylogenetic, or functional) distance matrix, with an extra parameter to adjust for the effect of closely related species, therefore allowing controlling the impact of PCR and sequencing errors. Given these features, Taberlet et al., 2018 highlighted the potential of the Hill number as an index of biodiversity fine-tuned for eDNA data.

1.4.5 Partitioning samples to find hidden structure

In many biodiversity studies, the main objective is to find hidden structure of species (taxon) assemblages in multiple different sites by comparison among them along some spatial/temporal or environmental gradients. Such comparisons can be performed using two major types of techniques: clustering and ordination (Legendre & Legendre, 2012).

Cluster analysis is the partitioning the collection of objects (species assemblages) under study. The purpose of cluster analysis is to find discontinuities in a dataset. There is a plethora of clustering methods, each of which adapts to some specific data types and meets specific need. Generally speaking, the main purpose of performing a clustering is to bring

out some hidden structures in data, and it is the user who decides whether these patterns are ecologically worth interpreting (Daniel Borcard, Gillet, & Legendre, 2011). In sedimentary DNA metabarcoding studies, the stratigraphically ordered nature of samples often poses a temporal constrain of the clustering. A traditional solution widely used in pollen analysis is the Constrained Incremental Sums of Squares (CONISS) algorithm (Grimm, 1987), which is also applied to sedimentary DNA data in a few studies (e.g. Zimmermann et al., 2016). This algorithm works by iteratively merging the two stratigraphically neighboring clusters that give least increase in within-cluster variance (compared to the sum of the dispersions of the two original clusters), which is calculated as sum of squares of a chosen dissimilarity coefficient (e.g. a distance measure). The result of a CONISS analysis is a hierarchical structure of sample clusters, which can be represented as a tree (i.e. a dendrogram). Like most clustering methods, CONISS does not provide a test to determine the optimal number of clusters (or zones). This task is usually done by comparing the variance reduction by each level of clustering to those expected from a randomly splitting with the same number of zones [e.g. those resulted from a broken-stick model, (Bennett, 1996b)].

A more recent method, multivariate regression tree (MRT) analysis (De'ath, 2002), allows clustering samples under the control of one or multiple, quantitative or categorical, explanatory variables. As the output of CONISS, the resulting MRT is also a hierarchical tree, whose "leaves" are subsets of samples chosen to minimize the within-group variance, but with each consecutive partition being defined by a threshold value (for the case of a quantitative explanatory variable) or a state (for the case of a categorical variable) (Daniel Borcard et al., 2011). A cross-validation process can be applied to determine to the most reasonable partition size, i.e. to decide at which level to retain the clustering. The process is a prediction-oriented method that randomly splits the set of samples into a training set to construct a MRT, and a much smaller testing set to validate the predictive power of the constructed MRT. The optimal partition size is defined as the level that minimizes the predictive error. Alternatively, a small partition size can be retained, if its predictive errors is one standard error above the minimal predictive error, which is call the 1-SE rule.

The MRT analysis is an efficient tool to partition stratigraphic samples, and has been applied to sedimentary DNA data (Ficetola, Poulenard, et al., 2018). In **Chapter 3**, we also use MRTs to find distinct phases of vegetation development based on sedimentary DNA. Besides, although the MRT analysis is originally a prediction-oriented machine-learning technique (Daniel Borcard et al., 2011), it can also serve as a explanatory tools to investigate biodiversity-environment relationships (see *1.4.7*).

1.4.6 Finding main trends by ordination

In contrast to cluster analysis, which seeks discontinuities in a dataset, ordination analysis aims to find main trends in the form of continuous gradients (Daniel Borcard et al., 2011). Indeed, biodiversity data such as a species abundance-site matrix is often multidimensional and therefore difficult to visualize. Ordination methods allow extracting the most "important" dimensions, in terms of e.g. variance they carry, for visualization and further analysis. As for clustering analysis, plenty of ordination methods are currently available, adapted for different data type and for different assumptions on the data. For instance, principle component analysis (PCA) works on quantitative species-site matrices, with the assumption that species' responses to environmental gradients are linear, whereas correspondence analysis (CA) works on non-negative, frequency-like matrices, and assumes unimodal responses to environmental gradients. Both PCA and CA, together with other eigenvalue-based methods such as principle coordinates analysis (PCoA), seek to represent a matrix in a series of axes ordered by their "importance", at the same time preserving the distances of a particular type (e.g. Euclidean in the case of PCA) among objects. In most cases, raw data must be properly transformed, in order to prevent the double-zero issue [i.e. the simultaneous absence of a taxon in both samples being considered a sign of similarity between them, but actually it should not (Legendre & Legendre, 2012)] from misleading the method. For eDNA data, a potentially favorable choice is the Euclidean distance computed on Box-Cox-chord-transformed data (Legendre & Borcard, 2018). This approach is a generalization of several traditional data transformations designed to simultaneously solve the double-zero problem and improve the multinormality of data (Legendre & Legendre, 2012). With a parameter to adjust for

different degree of data skewness (Legendre & Borcard, 2018), this approach would provide a flexible way to compare community compositions recorded in eDNA in avoiding the double-zero issue.

In contrast, non-metric multidimensional scaling (nMDS) works by an optimization process that tries to place objects in a space of a given number of dimensions, in a way that minimizes the total squared differences between the between-object distances in this space and those in the original space. In other words, nMDS does not preserve any particular between-object distance, and finds a configuration of points (objects) in a space of a lower number of dimensions with least changes possible in the distance relationships among points (Kruskal, 1964). Therefore, nMDS works on resemblance (distance, dissimilarity or similarity) matrices and does not assume any response mode. Consequently, nMDS is suitable for analyses that aim to restore species composition, rather than to find environmental gradients [see also (Minchin, 1987), which suggests that nMDS is more robust for indirect gradient analysis than the eigenvalue-based methods].

Given the flexibility and robustness of nMDS, there is no wonder why many sedimentary DNA metabarcoding studies chose it as the first step to characterize the similarity of communities (e.g. Ficetola et al., 2018; Pansu, De Danieli, et al., 2015; Pansu, Winkworth, et al., 2015). Most importantly, the relationship between DNA data (represented by number of reads or of number of positive replicates) and taxon abundance remain unclear (see *1.4.3*), and assumptions such as multivariate normality, continuity are rarely verified in metabarcoding studies, thus favoring nMDS. Despite its advantages, a successful nMDS analysis demands some extra cares. First, an adequate similarity or dissimilarity coefficient is needed to transform the raw matrix, and the choice of such coefficient is not evident. In some DNA metabarcoding applications, the Bray-Curtis dissimilarity is preferred, since it avoid the double-zero problem, and can deal with both presence/absence and abundance data (Pansu, De Danieli, et al., 2015; Pansu, Giguet-Covex, et al., 2015), but some other options are available. For instance, the Renkonen (dis)similarity is a preferred index when dealing with relative abundances, because of its density invariance (Jost, Chao, 2011). This feature may be favorable for comparing taxon compositions of sediment samples with

different amounts of DNA, as different degrees of DNA degradation that create such disparity are common in sediment samples. Second, there is no simple rule to determine in advance how many dimensions need to be retained in nMDS. Usually two to three axes are enough, but it is recommended to check the stress (the penalty function minimized by the nMDS algorithm) in several alternatives in order to help deciding. Third, although its goodness-of-fit can be assess by the above-mentioned stress, nMDS provides no intuitive measure of the representativeness of its axes upon the original dataset, as PCA does with the variance explained by each of its axis (represented by the respective eigenvalues). Therefore, it requires a different strategy to interpret the axes of an nMDS result. Finally, the axes of an initial nMDS result do not lie in parallel with any environmental gradient in most case. If needed, one can rotate the axes so that the first axis is parallel to an external variable without changing the configuration of points.

1.4.7 Testing the effects of environmental variables on community composition

The above-mentioned (unconstrained) ordination methods do not provide statistical tests of factors related to community variation, they only serve to represent the major features of the data with a reduced number of dimensions. To test specific hypotheses statistically, other methods are needed. To test the impacts of environmental factors on community composition, canonical analysis (or constrained ordination) is a popular choice. Canonical analysis is a class of statistical methods that generally combine the analysis of eigenvalues and eigenvectors (as in PCA, PCoA, CA etc., see *1.4.6*) and regression (Legendre & Legendre, 2012). It is called "canonical" because the methods rely on the "canonical" form of matrix, i.e. the matrix of eigenvalues. Canonical analysis produces ordinations vectors to explain that are constrained to be related in a certain way to the matrix of explanatory factors, thus it is also referred to as constrained ordination analysis. The way the ordinations are related to the matrix of explanatory variables is the defining feature of different methods in canonical analysis. In the case of redundancy analysis (RDA), an example of canonical analysis widely used in dealing with eDNA data, the ordination vectors are linear combinations of the explanatory variables. RDA can therefore be considered as an extension of PCA. On the other hand, in canonical correspondence

analysis (CCA), the ordination vectors are modeled in a unimodal relationship with the explanatory variables. Therefore CCA is the canonical counterpart of CA (Daniel Borcard et al., 2011). Usually, the significance of the ordination vectors resulted from a canonical analysis cannot be tested with classical parametric tests due to the non-normality of data, but it is possible to test it through permutation instead, by randomly permuting certain elements of the data to generate a distribution of the chosen statistic under the null hypothesis. Furthermore, the discussion on data transformation prior to ordination analysis (*1.4.6*) also applies to the canonical ordination here discussed. The RDA approach has provided valuable insights to the environment-community relationships in the past through sedimentary DNA. For example, using RDA on sedimentary DNA data, Pansu, Giguet-Covex, et al., 2015 showed that the plant community changes in an alpine lake area through the past six thousand years were mainly driven by livestock farming, and not by temperature changes. In a similar manner, Ficetola et al., 2018 demonstrated that the invasion of rabbit was responsible for a major change in plant species composition in one of the remote Kerguelen Islands.

Ecological phenomena often involve multiple tangled links. It is therefore crucial to resolve their confounding effects in analysis. In linear regression, it is possible to estimate the effects of some explanatory variables, at the same time controlling for the effects of some covariables. Similar approaches exist in canonical analysis, e.g. partial RDA and partial CCA (Legendre & Legendre, 2012). Especially, eDNA data are often spatially or temporally structured, so that part of the variations in such data may be originated from intrinsic (or neutral) spatial/temporal processes, of which the effects may overlap with those caused by "real" environmental factors, i.e. resulted from niche-based processes (Legendre & Legendre, 2012). Using partial constrained ordinations, it is possible to separate (or "partial out") the effects of the spatial/temporal components of variation from the ones caused by environmental gradients (Borcard, Legendre, & Drapeau, 1992).

The spatial and temporal structures of community composition variation also imply that ecological processes can happen at different scales. In the case of paleoecology, it is widely acknowledged that even the same environmental factor can have different effects on ecosystem in different time scale (Mills et al., 2016). Recent developments in spatial/temporal analyses provide a series of methods to assess the multiscale

spatial/temporal patterns in multivariate data. The Moran's eigenvector maps (MEMs) [also known as principal coordinate analysis of neighbor matrices (PCNM) (Dray, Legendre, & Peres-Neto, 2006)] is a family of to assess non-directional patterns. MEMs work by extracting eigenvectors that maximize autocorrelation, out of the spatial/temporal weighting matrix that summarizes the spatial/temporal relationships between samples. Directional patterns across space/time can also be assessed, with asymmetric eigenvector maps (AEMs) modeling, the directional counterpart of MEMs (Blanchet, Legendre, & Borcard, 2008). In practice, the resulting eigenvectors from both MEMs and AEMs can be used as explanatory variables in a constrained ordination analysis, in which the most relevant eigenvectors can be selected through forward selection based on their partial R^2 (Legendre & Gauthier, 2014). Furthermore, graphic representations can be drawn to highlight the essential scales of spatial/temporal variations using the multi-scale pattern analysis (MSPA), which applies RDA to the MEMs [as in (Legendre & Gauthier, 2014)] and then performs a variation of PCA (called %PCA) on the matrix of resulting centered R^2 (Jombart et al., 2009).

The methods presented so far in this section are essentially variations of multivariate analysis of variance (MANOVA) (Daniel Borcard et al., 2011), which partitions multivariate variance into components that are attributable to different explanatory variables. This approach by itself is not able to account for all potential response modes of ecosystem to environmental variables. For example, plant community compositional variability can be altered as a result of change in disturbance regime (Collins, 2000). In such cases, environmental changes lead to changes in the variance of some properties of the system, in addition to the properties themselves. Tests for homogeneity of variance such as Levene's test (Levene, 1961) can be applied to evaluate such effects for univariate response data, given that the explanatory variables divide the samples into several distinct groups. A multivariate analogue of these tests widely used in ecological studies is the test for homogeneity of multivariate dispersions (Anderson, 2006). In **Chapter 3**, we will use this test to demonstrate the destabilizing effect of livestock farming on plant community composition in an alpine area, based on sedimentary DNA metabarcoding data recording plant and mammal DNA over the last six thousand years.

1.4.8 Testing complex causal relationships

Ecosystems function in complex ways, involving multiple biotic/abiotic factors simultaneously at work, which are woven into a network of causality. Ecological explanations are therefore often both inherently multivariate and causal in nature (Shipley, 2000b). Long-term ecological records such as sedimentary DNA, combined with other proxies recording environmental conditions [e.g. chironomid fly records to reconstruct summer temperature in lake water (Heiri, Lotter, Hausmann, & Kienast, 2003; Korhola, Vasko, Toivonen, & Olander, 2002)], provide excellent opportunities to investigate such complex causal network. Instead of testing particular causal relationships one-by-one, it is possible to test a hypothetical causal network as a whole with structural equation modeling [SEM, (Grace, 2006; Shipley, 2000b)]. Treating causal relationship network as a whole brings benefits to the interpretation of data. Most notably it allows to establish causal relationships between variables and not merely correlations. In contrast to the traditional tenet "correlation does not imply causation", the principle of SEM lies in finding causations out of correlation. This is possible due to the fact that the ambiguous causality direction [either A causes B, or B causes A, or both A and B are caused by an unobserved C? (Shipley, 2000b)] in a correlation can be resolved by testing all the possibilities based on a priori knowledge of the system in question and the observational data at hand, using statistical techniques originally developed by Pearl (Pearl, 1988). In fact, hypothesized causal relationships can be represented as paths in a directed acyclic (or loop-free) graph, in which variables are represented as vertices and can appear simultaneously as predictors and responses (Lefcheck, 2016). By analyzing the structure of the graph, one can formulate a set of claims on the conditional independence among variables. Testing these claims provides a way to infer if there is no missing relationships among unconnected variables [Shipley's d-separation test and its extensions (Shipley, 2000a, 2009)]. This approach allows for the fitting of wide range of distributions and sampling designs, with ordinary linear regression models or generalized mixed models (Lefcheck, 2016; Shipley, 2009).

In **Chapter 4**, we present a case study applying SEM to a multiproxy paleorecord dataset (including sedimentary DNA metabarcoding data), to resolve the complex relationships among climate, land-use change, plant community development, and soil erosion. This case

demonstrates the potential of using multiproxy long-term records and advanced statistical methods to answer important ecological questions that are crucial to the wellbeing of humanity in the Anthropocene.

1.5 Reference

Ait Baamrane, M. A., Shehzad, W., Ouhammou, A., Abbad, A., Naimi, M., Coissac, E., … Znari, M. (2012). Assessment of the food habits of the Moroccan dorcas gazelle in M'Sabih Talaa, west central Morocco, using the trnL approach. PLoS ONE, 7(4). http://doi.org/10.1371/journal.pone.0035643

Alsos, I. G., Lammers, Y., Yoccoz, N. G., & Edwards, M. E. (2018). Metabarcoding lake sediments: taphonomy and representation of contemporary vegetation in environmental DNA (eDNA) records, (1).

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. Plos One, 13(4), e0195403. http://doi.org/10.1371/journal.pone.0195403

Alsos, I. G., Sjörgren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., … Pedersen, M. W. (2016). Sedimentary ancient DNA from Lake Skartjorna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene, 26(4), 627–642. http://doi.org/10.1177/0959683615612563

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62(1), 245–253. http://doi.org/10.1111/j.1541-0420.2005.00440.x

Bailey, L. L., Simons, T. R., & Pollock, K. H. (2004). Estimating site occupancy and species detection probability parameters for terrestrial salamanders. Ecological Applications, 14(3), 692–702. http://doi.org/10.1890/03-5012

Bellemain, E., Davey, M. L., Kauserud, H., Epp, L. S., Boessenkool, S., Coissac, E., … Brochmann, C. (2013). Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. Environmental Microbiology, 15(4), 1176–1189. http://doi.org/10.1111/1462-2920.12020

Bennett, K. D. (1996). Determination of the number of zones in a biostratigraphical sequence. New Phytologist, 132(1), 155–170. http://doi.org/10.1111/j.1469-8137.1996.tb04521.x

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Modelling directional spatial processes in ecological data. Ecological Modelling, 215(4), 325–336. http://doi.org/10.1016/j.ecolmodel.2008.04.001

Borcard, D., Gillet, F., & Legendre, P. (2011). Numerical ecology with R (Use R!). Springer. http://doi.org/10.1007/978-0-387-78171-6

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. Ecology, 73(3), 1045–1055. http://doi.org/10.2307/1940179

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. http://doi.org/10.1111/1755-0998.12428

Brose, U., Hillebrand, H., & Brose, U. (2016). Biodiversity and ecosystem functioning in dynamic landscapes.

Capo, E., Debroas, D., Arnaud, F., Guillemot, T., Bichet, V., Millet, L., … Isabelle Domaizon. (2016). Long-term dynamics in microbial eukaryotes communities: a paleolimnological view based on sedimentary DNA. Molecular Ecology. http://doi.org/10.1111/mec.13893

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. Annual Review of Ecology, Evolution, and Systematics, 45(1), 297–324. http://doi.org/10.1146/annurev-ecolsys-120213-091540

Chen, Y., Ye, W., Zhang, Y., & Xu, Y. (2015). High speed BLASTN: An accelerated MegaBLAST search tool. Nucleic Acids Research, 43(16), 7762–7768. http://doi.org/10.1093/nar/gkv784

Coissac, E., Hollingsworth, P. M., Lavergne, S., & Taberlet, P. (2016). From barcodes to genomes: extending the concept of DNA barcoding. Molecular Ecology, n/a-n/a. http://doi.org/10.1111/mec.13549

Collen, B., Loh, J., Whitmee, S., McRae, L., Amin, R., & Baillie, J. E. M. (2009). Monitoring Change in Vertebrate Abundance: the Living Planet Index. Conservation Biology, 23(2), 317–327. http://doi.org/10.1111/j.1523-1739.2008.01117.x

Collins, S. L. (2000). Disturbance Frequency and Community Stability in Native Tallgrass Prairie. The American Naturalist, 155(3), 311–325. http://doi.org/10.1086/303326

Crutzen, P. J. (2002). Geology of mankind. Nature, 415(January), 2002.

De'ath, G. (2002). Multivariate Regression Tree: A New Technique for Modeling Species–Environment Relationships. Ecology, 83(4), 1105–1117. http://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2

Dearing, J. A., Battarbee, R. W., Dikau, R., Larocque, I., & Oldfield, F. (2006). Human-environment interactions: Learning from the past. Regional Environmental Change, 6(1–2), 1–16. http://doi.org/10.1007/s10113-005-0011-8

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog Lithobates catesbeianus. Journal of Applied Ecology, 49(4), 953–959. http://doi.org/10.1111/j.1365-2664.2012.02171.x

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16(2), 129–138. http://doi.org/10.1111/j.1466-8238.2006.00279.x

Dougherty, M. M., Larson, E. R., Renshaw, M. A., Gantz, C. A., Egan, S. P., Erickson, D. M., & Lodge, D. M. (2016). Environmental DNA (eDNA) detects the invasive rusty crayfish Orconectes rusticus at low abundances. Journal of Applied Ecology, 53(3), 722–732. http://doi.org/10.1111/1365-2664.12621

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196, 483–493. http://doi.org/10.1016/j.ecolmodel.2006.02.015

Dunker, K. K. J., Sepulveda, A. J. A. A. J. A., Massengill, R. L. R., Olsen, J. J. B. J., Russ, O. L., Wenburg, J. J. K. J., … Shea, C. (2016). Potential of Environmental DNA to Evaluate Northern Pike (Esox lucius) Eradication Efforts : An Experimental Test and Case Study. PLoS ONE, 11(9), 1–21. http://doi.org/10.5061/dryad.16m53.Funding

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS ONE, 10(7), 1–16. http://doi.org/10.1371/journal.pone.0130324

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 1. http://doi.org/10.1186/1471-2164-11-434

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., … Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543–556. http://doi.org/10.1111/1755-0998.12338

Ficetola, G. F., Poulenard, J., Sabatier, D., Messager, E., Gielly, L., Leloup, A., … Arnaud, F. (2018). DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Science Advances. Science Advances. http://doi.org/10.1126/sciadv.aar4292

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources, 16(3), 604–607. http://doi.org/10.1111/1755-0998.12508

Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. Molecular Ecology Resources, 16(3), 641–654. http://doi.org/10.1111/1755-0998.12483

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., … Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications, 5, 3211. http://doi.org/10.1038/ncomms4211

Glikson, A. Y., & Groves, C. (2016). Climate, Fire and Human Evolution: The Deep Time Dimensions of the Anthropocene. http://doi.org/10.1007/978-3-319-22512-8

Grace, J. B. (2006). Structural equation modeling and natural systems. Cambridge University Press.

Grimm, E. C. (1987). Constrained Cluster Analysis By the Method of Incremental Sum of Squares. Computers and Geosciences, 13(1), 13–35. http://doi.org/10.1016/0098-3004(87)90022-7

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods in Ecology and Evolution, 8(9), 1081–1091. http://doi.org/10.1111/2041-210X.12743

Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, 68(4), 669–685. http://doi.org/10.1128/MBR.68.4.669–685.2004

Heiri, O., Lotter, A. F., Hausmann, S., & Kienast, F. (2003). A chironomid-based Holocene summer air temperature reconstruction from the Swiss Alps. The Holocene, 13(4), 477–484. http://doi.org/10.1191/0959683603hl640ft

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Nature Biotechnology, 11(9), 1026. http://doi.org/10.1038/ng0293-165

Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von, & Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research, 29(23), 4793–4799. http://doi.org/10.1093/nar/29.23.4793

Ives, A. R., & Zhu, J. (2011). Statistics for Correlated Data : Phylogenies , Space , and Time. Ecological Applications, 16(1), 20–32. Retrieved from //000236510100004

J. B. Kruskal. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29(1), 1–27.

Jombart, T., Dray, S., & Dufour, A. B. (2009). Finding essential scales of spatial variation in ecological data: A multivariate approach. Ecography, 32(1), 161–168. http://doi.org/10.1111/j.1600-0587.2008.05567.x

Jørgensen, T., Haile, J., Möller, P. E. R., Andreev, A., Boessenkool, S., Rasmussen, M., … Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Molecular Ecology, 21(8), 1989–2003. http://doi.org/10.1111/j.1365-294X.2011.05287.x

Jost, L., Chao, A., & Chazdon, R. L. (2011). Compositional similarity and beta diversity. Biological Diversity: Frontiers in Measurement and Assessment, 66–84.

Kéry, M. (2010). Introduction to WinBUGS for ecologists. Academic, Burlington.

Korhola, A., Vasko, K., Toivonen, H. T. T., & Olander, H. (2002). Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling. Quaternary Science Reviews, 21(16–17), 1841–1860. http://doi.org/10.1016/S0277-3791(02)00003-3

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. Molecular Ecology Resources, 16(3), 673–685. http://doi.org/10.1111/1755-0998.12486

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods in Ecology and Evolution, 7(5), 573–579. http://doi.org/10.1111/2041-210X.12512

Legendre, P., & Borcard, D. (2018). Box-Cox-chord transformations for community composition data prior to beta diversity analysis. Ecography. http://doi.org/10.1111/ecog.03498

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space − time analysis of community composition data. Proceedings of the Royal Society of London B: Biological Sciences, 281(1778), 20132728. http://doi.org/http://dx.doi.org/10.1098/rspb.2013.2728

Legendre, P., & Legendre, L. F. J. (2012). Numerical ecology (3rd ed.). Amsterdam: Elsevier.

Levene, H. (1961). Robust tests for equality of variances. Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling, 279–292.

Loreau, M., & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: A synthesis of underlying mechanisms. Ecology Letters, 16(SUPPL.1), 106–115. http://doi.org/10.1111/ele.12073

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J. P., Hector, A., … Wardle, D. A. (2001). Biodiversity and ecosystem functioning: Current knowledge and future challenges. Science, 294(5543), 804–808. http://doi.org/10.1126/science.1064088

Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew, J., & Langtimm, C. a. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. Ecology, 83(8), 2248–2255. http://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Marc, L., Babu, S., & Hamilton, K. (2005). Ecosystem Conditions and Human Well-being. Millenium Ecosystem Assessment. Ecosystems and HumanWell-Being: Current State and Trends, 123–164. http://doi.org/10.1079/PHN2003467

McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the anthropocene. Trends in Ecology and Evolution, 30(2), 104. http://doi.org/10.1016/j.tree.2014.11.006

Millenium Ecosystem Assessment. (2010). Ecosystems and Human Well-Being: Biodiversity Synthesis. Ecosystems, 285. http://doi.org/10.1057/9780230625600

Mills, K., Schillereff, D., Saulnier-Talbot, É., Gell, P., Anderson, N. J., Arnaud, F., … Mcgowan, S. (2016). Deciphering long‐term records of natural variability and human impact as recorded in lake sediments : a palaeolimnological puzzle. Wiley Interdisciplinary Reviews, 4(2), e1195. http://doi.org/10.1002/wat2.1195

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio, 69(1–3), 89–107. http://doi.org/10.1007/BF00038690

Muha, T. P., Rodríguez-Rey, M., Rolla, M., & Tricarico, E. (2017). Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. Frontiers in Ecology and Evolution, 5(December). http://doi.org/10.3389/fevo.2017.00158

Nichols, R. V, Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., … Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources, (February). http://doi.org/10.1111/1755-0998.12895

Olajos, F., Bokma, F., Bartels, P., Myrstener, E., Rydberg, J., Öhlund, G., … Englund, G. (2017). Estimating species colonization dates using DNA in lake sediment. Methods in Ecology and Evolution. http://doi.org/doi: 10.1111/2041-210X.12890

Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J. M., Gielly, L., Cordonnier, T., … C??cillon, L. (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA. Soil Biology and Biochemistry, 83, 100–105. http://doi.org/10.1016/j.soilbio.2015.01.004

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. Molecular Ecology, 24(7), 1485–1498. http://doi.org/10.1111/mec.13136

Pansu, J., Winkworth, R. C., Hennion, F., Gielly, L., Taberlet, P., & Choler, P. (2015). Long-lasting modification of soil fungal diversity associated with the introduction of rabbits to a remote sub-Antarctic archipelago. Biology Letters, 11(9), 20150408. http://doi.org/10.1098/rsbl.2015.0408

Parducci, L., Jorgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., … Willerslev, E. (2012). Glacial survival of boreal trees in northern Scandinavia. Science, 335(6072), 1083–1086. http://doi.org/10.1126/science.1216043

Parducci, L., Jorgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., … Willerslev, E. (2012). Glacial Survival of Boreal Trees in Northern Scandinavia. Science, 335(6072), 1083–1086. http://doi.org/10.1126/science.1216043

Parducci, L., Matetovici, I., Fontana, S. L., Bennett, K. D., Suyama, Y., Haile, J., … Willerslev, E. (2013). Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Molecular Ecology, 22(13), 3511–3524. http://doi.org/10.1111/mec.12298

Parducci, L., Valiranta, M., Salonen, J. S., Ronkainen, T., Matetovici, I., Fontana, S. L., … Suyama, Y. (2015). Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. Phil. Trans. R. Soc. B, 370(1660), 20130382. http://doi.org/10.1098/rstb.2013.0382

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (Representation and Reasoning). San Mateo, CA: Morgan Kaufmann.

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., … Kjær, K. H. (2013). A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quaternary Science Reviews, 75, 161–168. http://doi.org/10.1016/j.quascirev.2013.06.006

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., … others. (2016). Postglacial viability and colonization in North America's ice-free corridor. Nature, 537(7618), 45–49. http://doi.org/10.1038/nature19085

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. Molecular Ecology, 21(8), 1931–1950. http://doi.org/10.1111/j.1365-294X.2011.05403.x

Reid, W. V, Mooney, H. A., Cropper, A., Capistrano, D., Carpenter, S. R., Chopra, K., … others. (2005). Millennium Ecosystem Assessment. Ecosystems and human well-being: synthesis. World Resources Institute, Washington, DC.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research, 39(21), e145–e145. http://doi.org/10.1093/nar/gkr732

Ruddiman, W. F. (2013). The Anthropocene. Annual Review of Earth and Planetary Sciences, 41(1), 45–68. http://doi.org/10.1146/annurev-earth-050212-123944

Ryberg, M. (2015). Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. Molecular Ecology, 24(23), 5770–5777. http://doi.org/10.1111/mec.13444

Schloss, P. D., & Handelsman, J. (2005). Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 71(3), 1501–1506. http://doi.org/10.1128/AEM.71.3.1501–1506.2005

Schmidt, B. R., Kéry, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. Methods in Ecology and Evolution, 4(7), 646–653. http://doi.org/10.1111/2041-210X.12052

Seersholm, F. V., Pedersen, M. W., Søe, M. J., Taher, H., Mak, S. S. T., Ruter, A., … Hansen, A. J. (2016). DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4000 years ago -- supplementary. Nature Communications, 1(May), 1–16. http://doi.org/10.1038/ncomms13389

Shipley, B. (2000a). A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling, 7(2), 206–218. http://doi.org/10.1207/S15328007SEM0702_4

Shipley, B. (2000b). Cause and Correlation in Biology, (January 2000). http://doi.org/10.1017/CBO9780511605949

Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. Ecology, 90(2), 363–368. http://doi.org/10.1890/08-1034.1

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., … others. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Molecular Ecology Resources, 10(6), 1009–1018. http://doi.org/10.1111/j.1755-0998.2010.02855.x

Steffen, W., Crutzen, P. J., & McNeill, J. R. (2007). The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. AMBIO: A Journal of the Human Environment, 36(8), 614–621. http://doi.org/10.1579/0044-7447(2007)36[614:TAAHNO]2.0.CO;2

Steffen, W., Grinevald, J., Crutzen, P., & Mcneill, J. (2011). The anthropocene: Conceptual and historical perspectives. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369(1938), 842–867. http://doi.org/10.1098/rsta.2010.0327

Steffen, W., Persson, Å., Deutsch, L., Zalasiewicz, J., Williams, M., Richardson, K., … Svedin, U. (2011). The anthropocene: From global change to planetary stewardship. Ambio, 40(7), 739–761. http://doi.org/10.1007/s13280-011-0185-x

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21(8), 2045–2050.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14--e14. http://doi.org/10.1093/nar/gkl938

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., … Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology, 21(8), 1816–1820. http://doi.org/10.1111/j.1365-294X.2011.05317.x

Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., … Domaizon, I. (2018). Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution, 9(4), 1060–1069. http://doi.org/10.1111/2041-210X.12960

Vitousek, P. M., Mooney, H. a, Lubchenco, J., & Melillo, J. M. (1997). Human Domination of Earth' s Ecosystems. Science, 277(5325), 494–499. http://doi.org/10.1126/science.277.5325.494

Waters, C. N., Zalasiewicz, J., Summerhayes, C., Barnosky, A. D., Poirier, C., Gałuszka, A., … Wolfe, A. P. (2016). The Anthropocene is functionally and stratigraphically distinct from the Holocene. Science, 351(6269). http://doi.org/10.1126/science.aad2622

Weiß, C. L., Dannemann, M., Prüfer, K., & Burbano, H. A. (2015). Contesting the presence of wheat in the british isles 8,000 years ago by assessing ancient DNA authenticity from low- coverage data. ELife, 4(NOVEMBER2015), 1–10. http://doi.org/10.7554/eLife.10005

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., … others. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47–51. http://doi.org/10.1038/nature12921

Willoughby, J. R., Wijayawardena, B. K., Sundaram, M., Swihart, R. K., & DeWoody, J. A. (2016). The importance of including imperfect detection models in eDNA experimental

design. Molecular Ecology Resources, 16(4), 837–844. http://doi.org/10.1111/1755-0998.12531

WWF. (2016). Living planet report: risk and resilience in a new era. WWF International. http://doi.org/978-2-940529-40-7

Xu, C. C. Y., Yen, I. J., Bowman, D., & Turner, C. R. (2015). Spider Web DNA: A New Spin on Noninvasive Genetics of Predator and Prey. Plos One, 10(11), e0142503. http://doi.org/10.1371/journal.pone.0142503

Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. Molecular Ecology, 15(7), 1713–1731. http://doi.org/10.1111/j.1365-294X.2006.02882.x

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., … Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. Molecular Ecology, 21(15), 3647–3655. http://doi.org/10.1111/j.1365-294X.2012.05545.x

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., … Herzschuh, U. (2016). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). Biogeosciences Discussions, 14(3), 1–50. http://doi.org/10.5194/bg-2016-386

Chapter 2. Conditionally autoregressive models improve occupancy analyses of autocorrelated data: An example with environmental DNA (Article A)

Wentao Chen[1*] and Gentile Francesco Ficetola[1, 2]

[1] Univ. Grenoble Alpes, CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

[2] Department of Environmental Science and Policy, Università degli Studi di Milano. Via Celoria 26, 20133 Milano, Italy

* wentao.chen.ecol@gmail.com

Abstract

Site Occupancy-Detection Models (SODMs) are statistical models widely used for biodiversity surveys where imperfect detection of species occurs. For instance, SODMs are increasingly used to analyze environmental DNA (eDNA) data, taking into account the occurrence of both false-positive and false-negative errors. However, species occurrence data are often characterized by spatial and temporal autocorrelation, which might challenge the use of standard SODMs. Here we reviewed the literature of eDNA biodiversity surveys and found that most of studies do not take into account spatial or temporal autocorrelation. We then demonstrated how the analysis of data with spatial or temporal autocorrelation can be improved by using a conditionally autoregressive SODM, and show its application to environmental DNA data. We tested the autoregressive model on both simulated and real datasets, including chronosequences with different degrees of autocorrelation, and a spatial dataset on a virtual landscape. Analyses of simulated data showed that autoregressive SODMs perform better than traditional SODMs in the estimation of key parameters such as true/false positive rates, and show a better discrimination capacity (e.g. higher true skill statistics). The usefulness of autoregressive SODMs was particularly high in datasets with strong autocorrelation. When applied to real eDNA datasets (eDNA from lake sediment cores and freshwater), autoregressive SODM provided more precise estimation of true/false positive rates, resulting in more reasonable inference of occupancy states. Our results suggest that analyses of occurrence data, such as many applications of eDNA, can be largely improved by applying conditionally autoregressive specifications to SODMs.

Key-words: Sedimentary DNA; Species occupancy-detection model; Conditionally autoregressive model; Spatial autocorrelation; Temporal autocorrelation; True skill statistics

2.1 Introduction

Inferring species occurrence is fundamental in biodiversity studies. However, methods for species detection are often imperfect, as they can fail to detect present species, and can even be prone to false positives (Guillera-Arroita, 2017; Royle & Link, 2006). Failing to take into account these errors can lead to biased inference, therefore in the last years there was a development of Site Occupancy-Detection Models (SODMs) to deal with such issues. SODMs are able to estimate occupancy rates with imperfect detection (Mackenzie et al., 2002) and have been generalized to deal with false positives (Chambert, Miller, & Nichols, 2015; Miller et al., 2011; Royle & Link, 2006). In short, SODMs estimate the probability that a sample is occupied by a species given its detections on that sample, by considering the occupancy probability, true positive (TP)- and false positive (FP) probabilities over the whole sample set. SODM were originally developed to deal with the issues of traditional field surveys, but in the last years they are increasingly applied to a wide range of environmental data, including environmental DNA data (eDNA) (Dorazio & Erickson, 2018; Ficetola et al., 2015; Lahoz-Monfort, Guillera-Arroita, & Tingley, 2016; Schmidt, Kéry, Ursenbacher, Hyman, & Collins, 2013).

As with all the biodiversity data, eDNA generally does not detect all the present taxa (false negatives, or FNs), but is also subjected to FPs, which can arise through multiple processes such as contamination and PCR/sequencing errors (Ficetola et al., 2015). In many studies, scientific inference relies on the accurate assessment of occupancy state (presence or absence) of a given species in a particular sample, and both false presences and false absences can lead to severe issues (Ficetola, Taberlet, & Coissac, 2016). For example, in a study aiming at dating the introduction of alien species (e.g. Ficetola et al., 2018; Sjögren et al., 2017), FPs of the target species may increase the uncertainty of date estimates and even lead to misleading conclusions.

With traditional field survey data, SODMs can integrate calibration design (Chambert et al., 2015) or unbiased detections (Chambert et al., 2015; Miller et al., 2011) to account for detection errors. However, these approaches demand extra sampling efforts and may even be unfeasible for some eDNA applications, such as when eDNA is used to assess biodiversity in ancient samples (Capo, Debroas, Arnaud, & Domaizon, 2015; Epp et al., 2015; Giguet-Covex et al., 2014; Pansu, Giguet-Covex, et al., 2015). In such cases, FPs must be estimated from the detection dataset itself, even though negative and positive can provide some prior information on the frequency of these errors (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, Pedersen, et al., 2017).

SODMs are increasingly used to analyze ecological data, including eDNA (Pansu, Giguet-Covex, et al., 2015). These models generally assume independence of observations and thus do not take into account the spatial or temporal autocorrelation among samples. However, ecological variables commonly have some form of internal dependence, such as autocorrelation in space or time (Dormann, 2007; Roberts et al., 2017). When autocorrelation is present, samples that are nearby in space or in time often have similar occupancy because of both intrinsic processes (e.g. dispersion) and extrinsic mechanisms which in turn show autocorrelation (e.g. climate forcing) (Beale, Lennon, Yearsley, Brewer, & Elston, 2010; Ficetola, Manenti, De Bernardi, & Padoa-Schioppa, 2012; Wagner & Fortin, 2005). Autocorrelation is particularly important in ecological time series (Legendre & Gauthier, 2014; Legendre & Legendre, 2012), and ignoring the correlation between nearby samples violates a central assumption of many statistical methods, and may result in biased inferences.

Approaches exist to integrate autocorrelation into SODMs, which have been sometimes applied to traditional biodiversity surveys. To name a few, Sargeant, Sovada, Slivinski, & Johnson (2011) showed that Markov random fields (multidimensional extensions of Markov chains) can allow modeling spatial dependencies in estimating species distribution. Royle & Dorazio (2009) proposed the auto-logistic structure to account for temporal autocorrelation; Hines et al. (2010) used Markov process to model the probability of

presence of tiger along consecutive trail segments; while Aing, Halls, Oken, Dobrow, & Fieberg (2011) modeled the occupancy of river otter on snow track with an intrinsic conditional autoregressive (iCAR) term. However, despite the wide awareness of the potential biases introduced by the autocorrelation in spatially or temporally structured data, most eDNA surveys using SODMs to analyze such data did not explicitly consider autocorrelation.

Here we demonstrate the needs and benefits to consider spatial- and temporal autocorrelation in eDNA biodiversity surveys. We begin with an analysis of the literature, to identify the frequency of SODM analyses in the eDNA literature, and to assess to what extent conditionally autoregressive SODM could improve analyses. Then we show how SODMs with explicit autocorrelation structure can improve occupancy analyses on spatially or temporally structured eDNA data. To this end, we added a Conditionally Autoregressive (CAR) component into the Royle & Link (2006) SODM, to take into account that occupancy can be affected by autocorrelation. The Royle & Link (2006) model can be applied to survey data without calibration designs or unbiased detection data, and also takes into account the probability of FPs, which makes it suitable for general eDNA-based surveys (Ficetola et al., 2015; Lahoz-Monfort et al., 2016). CAR models had been demonstrated to be useful to model spatial autocorrelation in species distribution data, are easy to implement and interpret (Dormann, 2007), and using CAR or closely related restricted spatial regression models to account for autocorrelation in occupancy models can yield satisfying results (Aing et al., 2011; Johnson, Conn, Hooten, Ray, & Pond, 2013). Our approach can be particularly important in eDNA studies, where contamination can occur (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, Pedersen, et al., 2017), and at the same time TP rates can be low (Epp et al., 2015), as a result of low DNA content in samples, nevertheless similar issues also apply to other approaches to biodiversity assessment. We assessed the capacity of a CAR approach to improve the performance of SODM models and to better estimate true occupancy. We tested and compared the CAR-SODM with the original Royle & Link (2006) SODM by applying them to 1) simulated chronosequences of occupancy data, 2) simulated spatial samples and 3) true eDNA datasets from the literature.

Our study identifies the cases in which the CAR-SODMs perform better than the original SODMs, and provides easy-to-follow instructions to implement CAR-SODMs.

2.2 Materials and Methods

2.2.1 Analysis of the literature

To elucidate the issues of data autocorrelation in current literature of eDNA biodiversity studies that involved species occupancy modeling, we collected relevant studies from the Web of Science database (June 2018), using search keywords "DNA occupancy NEAR/2 model" and "DNA detection probability". Each resulting study was screened based on these criteria: 1) whether it reported empirical eDNA data; 2) whether the data showed a temporal or spatial structure; 3) whether studies applied SODMs on those data and 4) whether the SODMs used took autocorrelation into account if the data were temporally or spatially structured.

2.2.2 CAR-SODM

We adopted here an extended SODM based on the Royle & Link (2006) model, which can be specified in a Bayesian framework as:

$$z_i \sim \text{Bernoulli}(\psi_0) \quad (1)$$

$$y_i \sim \text{binomial}(z_i p_{11} + (1\text{-}z_i) p_{10}, K) \quad (2)$$

where $i = 1, 2, ..., n$, n being the total number of samples; ψ_0 is the overall occupancy probability in all samples; z_i the state of occupancy in sample i ($z_i = 1$ when the site is occupied, otherwise $z_i = 0$); y_i the number of positive detections in sample i; p_{11}, p_{10} the TP and FP probability, respectively. K is the number of replicated observations. For instance, in eDNA studies K can be the number of PCRs or of replicated DNA extractions

performed on a given sample (Ficetola et al., 2015). We note that the likelihood function of this model would have two equally high peaks, because of the symmetry between $z_i * p_{11}$ and $(1 - z_i) * p_{10}$ in (2). In other words, for a given dataset, a parameter set with $z_i = z'_i$, $p_{11} = p'_{11}$ and $p_{10} = p'_{10}$ has the same likelihood as another set with $z_i = 1\text{-}z'_i$, $p_{11} = p'_{10}$ and $p_{10} = p'_{11}$. Therefore, there are several solutions for a given dataset of y_i, so that additional data or restrictions on parameters are needed to get unambiguous estimation. As a result, for models with a TP/FP setting, additional data is needed to break such symmetry and obtain an unambiguous solution (Guillera-Arroita et al., 2017; Lahoz-Monfort et al., 2016; Royle & Link, 2006).

In the CAR-SODM here proposed, instead of treating ψ_0 as a constant parameter over the whole sample set, we note the occupancy probability at sample i as ψ_i, which can take different values for different is. To model the autocorrelation of ψ_i, we assume that ψ_i can be expressed by a baseline constant ψ_b and an autoregressive term φ_i drawn from a multivariate normal distribution (Jin, Carlin, & Banerjee, 2005):

$$\psi_i = \text{logistic}(\psi_b + \varphi_i) \quad (3)$$

$$\varphi_i \sim \text{multinormal}(0,\ \sigma^2\left[\text{Diag}(m) - \alpha W\right]^{-1}) \quad (4)$$

The term $\sigma^2[\text{Diag}(\mathbf{m}) - \alpha\mathbf{W}]^{-1}$ is the covariance matrix of the multivariate normal distribution. $\text{Diag}(\mathbf{m})$ is an n by n diagonal matrix with $\mathbf{m} = m_i$, i.e. the number of neighbors for sample i as its diagonal elements. $\mathbf{W}$ is the adjacency matrix: $W_{ij} = 1$ if $i \neq j$ and samples i and j are neighbors, otherwise $W_{ij} = 0$. The α parameter measures the degree of dependence among samples. $\alpha = 0$ corresponds to a model without temporal / spatial dependence. σ^2 is the variance parameter to the multivariate normal distribution (Jin et al., 2005; Wall, 2004).

2.2.3 Parameter estimation

One of our aims was using CAR-SODM to better estimate the true occupancy state on the basis of detection data, for instance by assessing the probability that a sample i is actually occupied, given the number of positive detections in that sample. This can be calculated from other parameters specified above (Lahoz-Monfort et al., 2016):

$$P_i(\text{occupied}|y) = P(z_i = 1|y_i = y) = \frac{\psi_i p_{11}^{y}(1-p_{11})^{K-y}}{\psi_i p_{11}^{y}(1-p_{11})^{K-y}+(1-\psi_i)p_{10}^{y}(1-p_{10})^{K-y}} \quad (5)$$

In this study, we compare the CAR-SODM with the Royle & Link (2006) model by observing their behavior and performance. We estimated all parameters involved in the CAR-SODM (p_{11}, p_{10}, φ_i, ψ_b, τ and α) and in the Royle & Link model (p_{11}, p_{10} and ψ_0) in a Bayesian framework. For both models, uniform priors were used for most parameters with boundaries adjusted to their proper definitions ((0, 1) for probabilities and (-1, 1) for the autocorrelation parameter α). For p_{10}, we used a uniform prior between 0 and 0.1 to reflect prior information obtained from positive controls (Pansu, Giguet-Covex, et al., 2015), and to break the symmetry between p_{11} and p_{10} in the likelihood function (Lahoz-Monfort et al., 2016). For σ^2, we used a $\text{gamma}(2, 1)$ prior. Sampling was performed by using Markov chain Monte Carlo (MCMC) sampling provided by the Stan statistical computation platform from the *RStan* interface (Stan Development Team, 2016). Each sampling run used three MCMC chains for 10,000 iterations, the first of which 5,000 were discarded as burn in. The convergence of chains was checked by the Gelman-Rubin statistic (Gelman, Rubin, Gelman, & Rubin, 1992). In all models, input data included number of replicates for each sample K, the observed number of positive detections y, the adjacency matrix **W** and the corresponding m_i. In fact, **W** can be specified according to the sampling scheme for each application.

2.2.4 Simulations: chronosequences

A first set of simulated data mimicked the occupancy data obtained by the analysis of temporal series, such as eDNA data obtained from the analysis of lake sediments. We generated the chronosequence simulation data based on a combination of the autoregressive

process (AR1) and the binomial process. In each simulation replicate, the sample numbers corresponded to their chronological order. First, a series of $X_t = -0.005 + \alpha_c X_{t-1} + \varepsilon_t$ was generated, where α_c is the autoregressive parameter and ε_t is an error term drawn from the standard normal distribution. The probability of occupancy ψ_t was obtained by applying the inverse-logit function to X_t. The detection count y_t was then generated as in (2). We additionally restricted the proportion of positive z_i between 5% and 95% to avoid unrealistic extreme cases. Different values of TP probability p_{11} (0.15, 0.25 and 0.4) and FP probability p_{10} (0.005, 0.015 and 0.03) were applied to the simulations. We kept p_{11} at relatively low levels because these are the most challenging cases of in real-world studies, and large p_{11}s (≥0.5) allow to obtain satisfying results even by applying the Royle & Link (2006) model (Ficetola et al., 2015; Lahoz-Monfort et al., 2016). Besides, small p_{11}s are common in both classical monitoring and in eDNA studies, especially in ancient DNA as a result of degradation (Ficetola, Romano, Salvidio, & Sindaco, 2018; Giguet-Covex et al., 2014; Pansu, Giguet-Covex, et al., 2015; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, Pedersen, et al., 2017). For each parameter scenario (p_{11}, p_{10}, K, α_c), 100 simulation replicates were performed. We chose three K levels (4, 8 and 12) to reflect common practices in eDNA analyses (Ficetola et al., 2015). Three levels of αc (0.0001, 0.5, 0.95) were used to represent realistic levels of temporal autocorrelation (see results), ranging from practically non-correlated ($\alpha_c = 0.0001$) to highly correlated data ($\alpha_c = 0.95$). Sample size S was set to 100 to represent typical ancient DNA datasets. Both the CAR-SORM and the Royle & Link (2006) models were applied to the simulation dataset. To build the adjacency matrix of the CAR-SODM, **W** was specified as such that consecutive samples were considered to be chronologically adjacent.

2.2.5 Simulations: spatially autocorrelated dataset

The datasets representing spatial data with autocorrelation were generated in two steps. First, we generated artificial distribution data for a virtual island using the SNOUTER dataset (Dormann et al., 2007). The original dataset consists of the X-Y coordinates in integers of more than 1000 sites and two uncorrelated environmental variables for each site:

"rain" and "djungle". To reflect real world applications, we used only a subset of sites for simulations (i.e., only the sites with both coordinates that were multiples of 3 were retained in the simulations, resulting in 120 sites). Following the authors' instructions, we calculated the presence probability p_i for a site i as

$$p_i = q_i + \sqrt{q_i(1-q_i)}\epsilon_i, \text{ where } q_i = \frac{e^{3-0.004*rain}}{1+e^{3-0.004*rain}}.$$

To produce the presence/absence data, we set $z_i = 1$ if $p_i \geq 0.5$ and $z_i = 0$ if $p_i < 0.5$. The resulting data are spatially correlated because "rain" is essentially determined by altitude with a rain-shadow in the east (the SNOUTER dataset itself is based on the digital elevation model of a volcano in New Zealand). Second, detection data y_i were generated by applying (2) to z_i. The same sets of parameter scenarios for p_{11}, p_{10} and K as in the chronosequence simulation section were used to generate the datasets. In the CAR-SODM, the adjacency matrix **W** was defined with a rook scheme.

2.2.6 Model performance comparison

We first assessed the ability of models to correctly infer p_{11} and p_{10}. Furthermore, in many applications of SODMs, users are interested in obtaining better information on the actual occupancy state of a sample (e.g. what is the probability that a given site is occupied, if I have only one detection over 8 sampling replicates?). We thus obtained model estimations of the conditional probability of occupancy of samples P_i (eqn. 5). To evaluate model performances in assessing the actual states of the simulation samples, we then calculated the Area Under the Curve of the Receiver Operating Plot (AUC) and the maximum True Skill Statistic (TSS). AUC is a threshold-independent measure for score-ranking models (Bradley, 1997) that reflects a model's overall classification accuracy. AUC values range from 0 to 1. Usually AUC values of 0.5–0.7 indicate low accuracy, values of 0.7–0.9 indicate useful applications and values of > 0.9 indicate high discrimination capacity (Swets, 1988). However, AUC has some limitations (Lobo, Jiménez-valverde, & Real, 2008), thus we used the maximum TSS as a complementary measure of performance. TSS

is a simple and intuitive, threshold-dependent measure of accuracy in predicting presence-absence (Allouche, Tsoar, & Kadmon, 2006). TSS values range from -1 to 1. TSS values of 1 indicate a perfect discrimination, while TSS values equal to or less than zero indicate a performance no better than a random model (Allouche et al., 2006). Because different statistical models can have different optimal prediction thresholds (Bradley, 1997), we maximized the TSS by varying the prediction threshold on the Receiver Operating Characteristic (ROC) curves for each model, in a similar manner to a previous study aimed at finding the optimal detection threshold for eDNA qPCR assays (Serrao, Reid, & Wilson, 2017).

2.2.7 Analysis of Empirical data

To demonstrate the usefulness of conditionally autoregressive models, we applied the model to pubished eDNA data available from the literature search. Specifically, we analyzed a subset of chronosequences in the Lake Anterne sedimentary ancient DNA dataset (Giguet-Covex et al., 2014; Pansu, Giguet-Covex, et al., 2015), as well as spatially structured survey data of animal species based on modern eDNA collected from water samples. We considered the studies obtained from the literature search that provided all the data required to apply the CAR-SODM model (i.e. geographic coordinates of sampling sites, and detection histories at sites) (Dougherty et al., 2016; Ostberg, Chase, Hayes, & Duda, 2018; Vörös, Márton, Schmidt, Gál, & Jelić, 2017). For the ancient DNA dataset, we chose six molecular operative taxonomic units (MOTUs) representing two animal (*Bos* and *Ovis*) and four plant taxa (*Achillea macrophylla*, *Alchemilla* MOTU, *Hypericum* MOTU and *Pinus* MOTU) from the dataset to evaluate the behavior of both models for taxa with different features. The Dougherty et al., (2016) dataset contains eDNA of rusty crayfish (*Orconectes rusticus*) collected from water samples from 12 lakes in Wisconsin, USA. We analyzed the whole eDNA dataset, considering multiple collection sites in the same lake as adjacent. The Vörös et al., (2017) dataset contains eDNA data of olm salamander (*Proteus anguinus*) from water samples collected from 15 cave systems in Croatia. We analyzed the eDNA data obtained with precipitation method, using the Gabriel graph criterion (Legendre & Legendre, 2012) to determine site connectivity because it is an approximation for an

irregular spatial distribution to the rook scheme for a regular grid (Bini et al., 2009). The Ostberg et al., (2018) dataset contains eDNA of pacific lamprey (*Entosphenus tridentatus*) and *Lampetra* spp from water samples collected from watersheds in Puget sound, USA. We analyzed the eDNA data of the 2015 spring survey, considering consecutive sites on the Puget sound watershed line as adjacent. Settings of the datasets (spatial/temporal structure, Sample size S and number of PCR replicates K) can be found in Table 2-2.

All simulations and analyses were performed in the R statistical programming environment (version 3.3.3) (R Core Team, 2017). ROC curves, AUC values were generated by using the *pROC* package (Robin et al., 2011). Spatial weights matrices used in generating the spatially autocorrelated dataset were created using the *spdep* package (Bivand, Hauke, & Kossowski, 2013; Bivand & Piras, 2015).

2.3 Results

2.3.1 Literature analysis

The Web of Science database returned 134 journal articles as search results. We discarded all articles that did not directly discuss eDNA and those that did not report eDNA data from field surveys, including pure simulation studies, theoretical analyses and those that focused on experimental tests on DNA detectability under controlled conditions. The remaining 51 articles, dating from 2008 to 2018, were analyzed in detail (Table 2-1).

		SODM			
Structure	**No SODM**	**Autocorrelation only**	**False positive only**	**Both**	**Neither**
Spatially structured	11	0	0	1(1)	18(17)

Temporally structured	1	0	0	1(0)	0
Both	0	0	0	0	3(2)
Neither	12	0	2(1)	0	2(0)
Total	24	0	2(1)	2(1)	23(19)

Table 2-1: Analysis of the literature using eDNA for biodiversity assessment. The table shows the number of studies that analyzed eDNA survey data with different typologies of autocorrelation (spatial/temporal). We report if papers analyzed data using site occupancy-detection models (SODM), considering autocorrelation, and taking into account false positives. Numbers in parentheses indicate the number of studies using SODMs to model the effects of covariates on occupancy or detection probability. See Appendix for the full list of screened papers.

Spatial or temporal structures were frequent, as they were present in 35/51 studies (69%). 27/51 studies (52%) used SODM to analyze eDNA data. However, just 4/27 (15%) of them took into account FPs in their SODMs, and only 2/23 (9.5%) of studies using SODMs on spatially or temporally structured data considered the autocorrelation issue (Table 1, see also Supporting Information Table for the list of the studies considered).

2.3.2 Simulations: chronosequences

With the non-autoregressive models, the values of TP probability p_{11} were slightly overestimated when the number of replications K was small and detection probability was low, while estimates of the autoregressive model were less biased and more precise (Figure 2-1). FP probability p_{10} and autocorrelation did not severely affect the estimation of p_{11} with both models. In scenarios with large K (≥ 8) or large p_{11} (≥ 0.25), both models estimated TP probability with high accuracy. The auto-regressive model also showed lower bias in the estimation of FP probability p_{10}, even though both approaches tended to overestimate p_{10}, unless a large number of replicates per sample was available and TP probability was high (Figure 2-2).

The model performance in assessing the actual state (presence / absence) of the samples was generally fair to good, with AUC values generally > 0.7, and TSS values often > 0.5 (Figure 2-3, Figure 2-S1). With both metrics, performance was better when TP probability was high, and when p_{10} was low (Figure 2-3, Figure 2-S1). In most scenarios, the autoregressive model outperformed the non-autoregressive one. CAR-SODM provided a particularly high improvement compared to the Royle & Link (2006) model when autocorrelation was strong, when TP probability was low and when a few replicates were performed. In scenarios with high autocorrelation and low replication level, the autoregressive model outperformed its non-autoregressive counterpart by more than 10% in terms of AUC, even with relatively high TP probability ($p_{11} = 0.4$). Both models showed excellent performance (AUC and TSS close to one) if many replicates were performed and TP probability was high (Figure 2-3, Figure 2-S1).

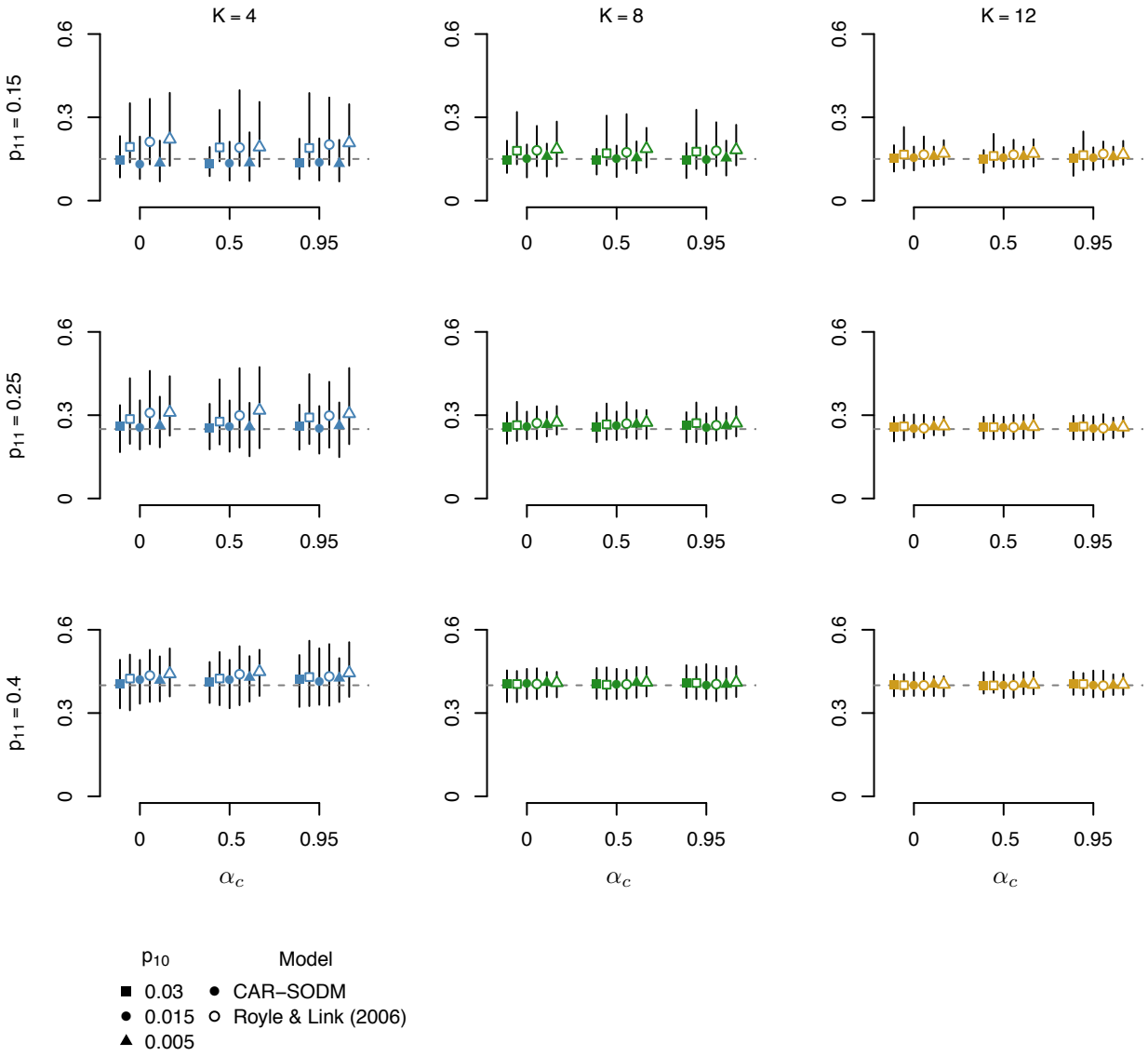


Figure 2-1: Estimations of true-positive (TP) probabilities p_{11} for the eDNA chronosequence simulated datasets by the CAR-SODM (full symbols) and by the Royle & Link (2006) model (open symbols). Results are arranged by scenarios of p_{11}, autocorrelation level α_c and false-positive (FP) probability p_{10}. Plots show medians from 100 simulations. Vertical lines cover the 2.5%-97.5% quantiles. Dashed lines indicate the actual p_{11} levels at 0.15 (top), 0.25 (middle) and 0.4 (bottom).

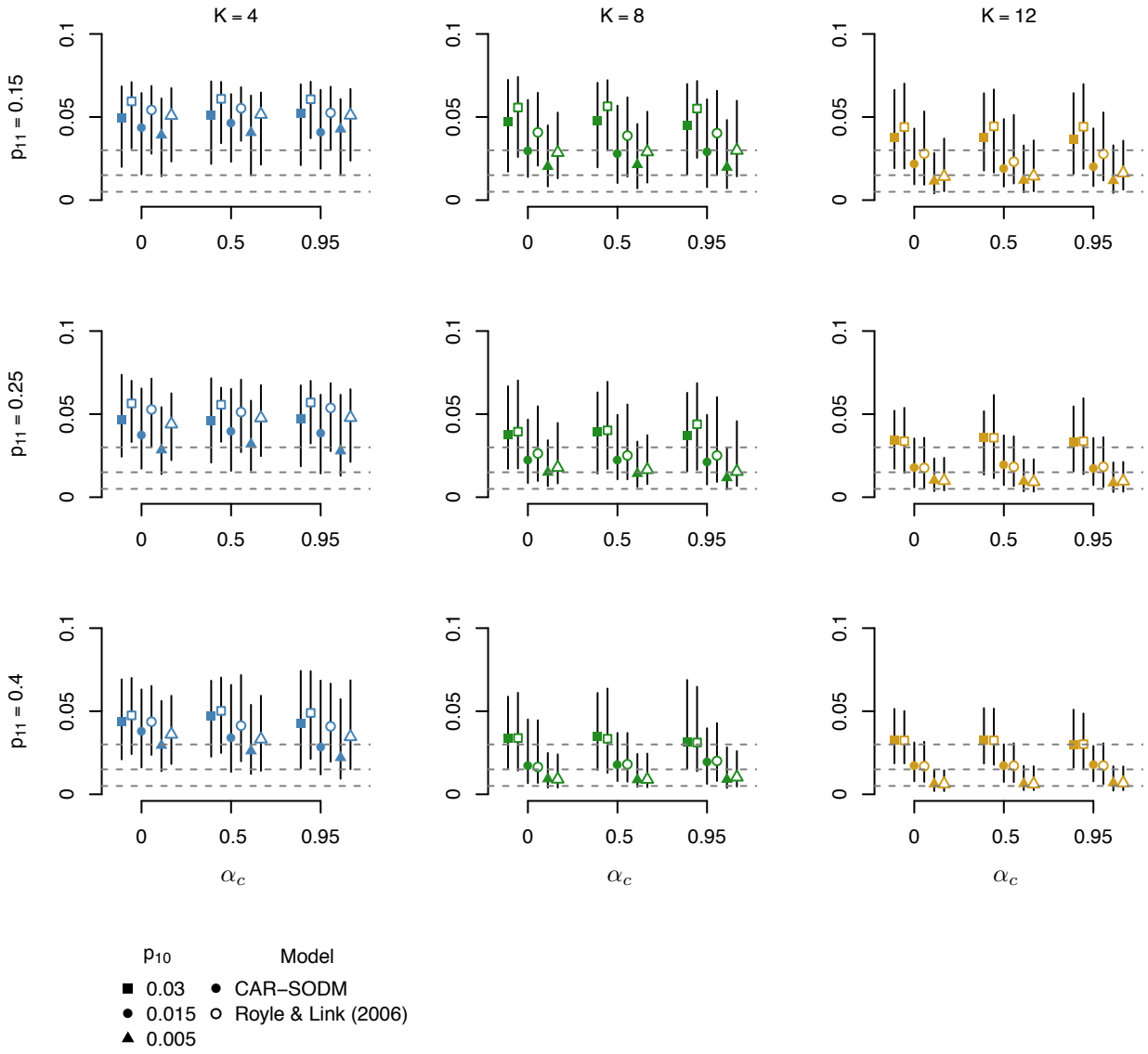


Figure 2-2: Estimations of false-positive (FP) probabilities p_{10} for the eDNA chronosequence simulated datasets by the CAR-SODM (full symbols) and by the Royle & Link (2006) model (open symbols). Results are arranged by scenarios of true-positive (TP) probability p_{11}, autocorrelation level α_c and (FP) probability p_{10}. Plots show medians from 100 simulations. Vertical lines cover the 2.5%-97.5% quantiles. Dashed lines indicate the actual p_{10} levels at 0.005, 0.015 and 0.03.

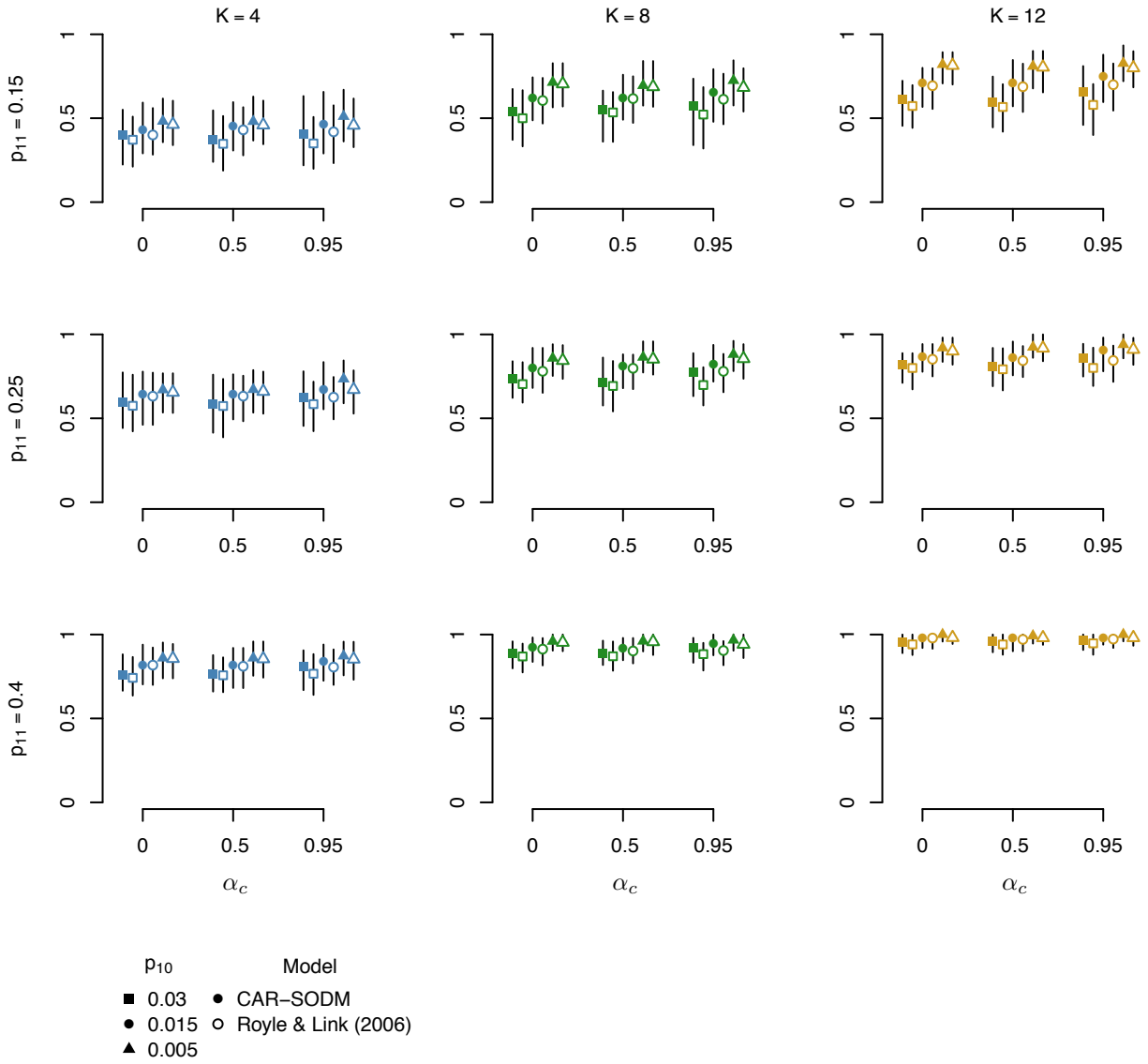


Figure 2-3: Maximum True Skill Statistic (TSS) for the CAR-SODM and the Royle & Link (2006) model, applied to the chronosequence simulated datasets. Results are arranged by scenarios of true-positive (TP) probability p_{11}, autocorrelation level α_c and false-positive (FP) probability p_{10}. Full symbols: CAR-SODM; open symbols: Royle & Link (2006). Plots show values calculated over 100 simulations. Vertical lines cover the 2.5%-97.5% quantiles.

2.3.3 Spatial data simulations

The analysis of data with spatial autocorrelation yielded results consistent with the ones of the chronosequence simulations. In comparison to the Royle & Link (2006) model, the CAR-SODM provided more accurate estimates of p_{11} and p_{10}, particularly in scenarios of low TP probability and with a few replicates (Figure 2-4). In terms of TSS and AUC, the autoregressive model outperformed the non-regressive one in most scenarios, except in scenarios of high p_{11}, K and low p_{10}, where both models reached excellent performance (Figure 2-4, Figure 2-S2).

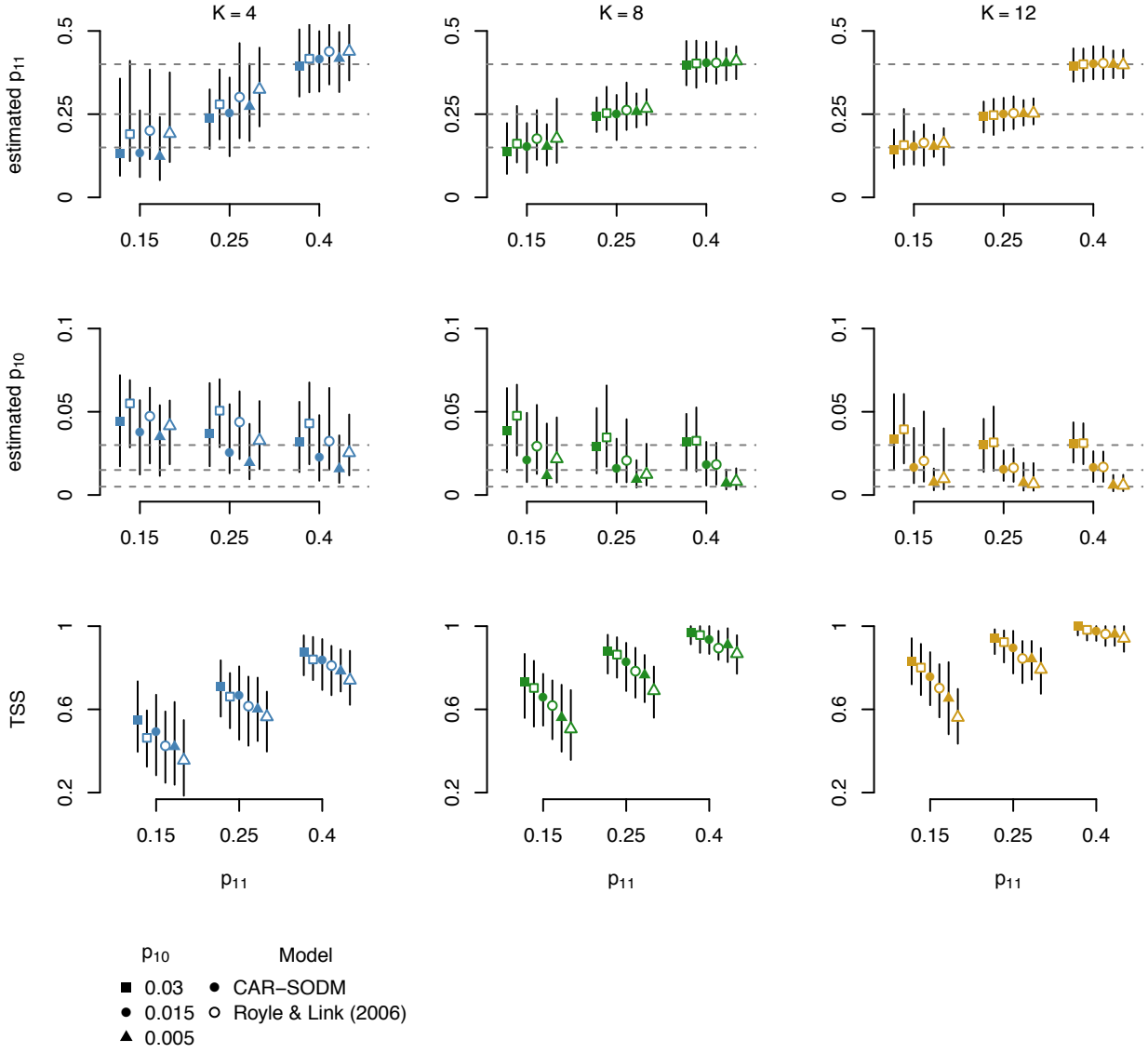


Figure 2-4: Estimations of true-positive (TP) probabilities p_{11} (top row) and of false-positive (FP) probability p_{10} (middle row), applied to the simulated dataset with spatial structure, analyzed with the CAR-SODM (full symbols) and by the Royle & Link (2006) model (open symbols). Plots show medians from 100 simulations. Vertical lines cover the 2.5%-97.5% quantiles. Bottom row: Maximum TSS for the CAR-SODM and the Royle & Link (2006) model. All results are arranged by scenarios of p_{11} and p_{10}. Dashed lines indicate p_{11} levels at 0.15, 0.25 and 0.4 (top row), and p_{10} levels at 0.005, 0.015 and 0.03 (middle row).

2.3.4 Analysis of empirical eDNA data

We then applied both the CAR-SODM and the Royle & Link (2006) model to multiple real-world datasets. When we analyzed the eDNA data from two mammal and four plant MOTUs from 44 samples of lake sediment (Giguet-Covex et al., 2014; Pansu, Giguet-Covex, et al., 2015), the CAR-SODM detected strong positive autocorrelation for all taxa (estimated $\alpha \geq 0.78$). The CAR-SODM tended to estimate lower values of both p_{10} and p_{11} than the Royle & Link (2006) model. Furthermore, highest probability density intervals (HPDIs) for corresponding parameters tended to be narrower for CAR-SODM, suggesting lower incertitude of estimates. CAR-SODM models were also applied to spatially structured datasets of modern eDNA collected from water samples. When we analyzed the dataset on crayfish eDNA (Dougherty et al., 2016), we obtained results analogous to those of the ancient DNA, with strong positive autocorrelation ($\alpha = 0.932$) and lower p_{10} and p_{11} estimated by the CAR-SODM than by the Royle & Link (2006) model. Conversely, the CAR-SODM did not detect significant autocorrelation in the olm salamander (Vörös et al., 2017) and in the lamprey (Ostberg et al., 2018) eDNA datasets (HPDIs of α included zero). In these two datasets, both models yielded nearly identical results in the estimates of p_{10} and p_{11} (Table 2-2).

Paper	Structure	S	K	Taxon	α CAR	ψ_b CAR	ψ_0 RL	p_{11} CAR	p_{11} RL	p_{10} CAR	p_{10} RL
(Giguet-Covex et al., 2014)	Temporal	44	8	*Bos* (cattle)	0.864	0.395	0.333	0.302	0.313	0.008	0.012
					(0.585,1.000)	(0.001, 0.887)	(0.180, 0.519)	(0.212, 0.398)	(0.208, 0.428)	(0.000, 0.024)	(0.000, 0.033)

	Temporal	44	8	*Ovis* (sheep)	0.780 (0.034,1.000)	0.396 (0.000, 0.905)	0.167 (0.031, 0.328)	0.261 (0.098, 0.434)	0.338 (0.118, 0.510)	0.012 (0.000, 0.028)	0.015 (0.000, 0.035)
(Pansu, Giguet-Covex, et al., 2015)	Temporal	44	8	*Achillea macrophylla*	0.986 (0.951,1.000)	0.401 (0.000, 0.903)	0.155 (0.042, 0.272)	0.410 (0.281, 0.54)	0.491 (0.315, 0.681)	0.010 (0.001, 0.023)	0.016 (0.001, 0.035)
	Temporal	44	8	*Alchemilla* MOTU	0.948 (0.817,1.000)	0.415 (0.000, 0.913)	0.285 (0.153, 0.424)	0.820 (0.700, 0.93)	0.871 (0.784, 0.958)	0.039 (0.002, 0.072)	0.053 (0.021, 0.088)
	Temporal	44	8	*Hypericum* MOTU	0.928 (0.736,1.000)	0.310 (0.000, 0.857)	0.225 (0.093, 0.380)	0.513 (0.376, 0.656)	0.518 (0.352, 0.671)	0.060 (0.031, 0.092)	0.061 (0.030, 0.095)
	Temporal	44	8	*Pinus* MOTU	0.787 (0.230,1.000)	0.307 (0.000, 0.829)	0.218 (0.062, 0.398)	0.38 (0.232, 0.529)	0.438 (0.264, 0.633)	0.042 (0.011, 0.077)	0.051 (0.017, 0.087)
(Dougherty et al., 2016)	Spatial	262	4	*Orconectes rusticus* (rusty crayfish)	0.932 (0.845, 0.996)	0.338 (0.000, 0.791)	0.331 (0.222, 0.433)	0.739 (0.647, 0.830)	0.829 (0.730, 0.924)	0.043 (0.010, 0.079)	0.073 (0.042, 0.100)
(Vörös et al., 2017)	Spatial	15	20	*Proteus anguinus* (olm salamander)	0.440 (-0.989, 0.978)	0.118 (0.003, 0.264)	0.319 (0.000, 0.876)	0.681 (0.491, 0.866)	0.682 (0.492, 0.86)	0.039 (0.018, 0.063)	0.039 (0.020, 0.067)
(Ostberg et al., 2018)	Spatial	18	12	*Entosphenus tridentatus* (pacific lamprey)	0.212 (-0.739, 0.999)	0.272 (0.000, 0.802)	0.209 (0.051, 0.387)	0.731 (0.529, 0.885)	0.740 (0.564, 0.891)	0.028 (0.000, 0.054)	0.030 (0.003, 0.058)
	Spatial	18	12	*Lampetra* spp	0.305 (-0.529, 0.998)	0.691 (0.204, 1.000)	0.7 (0.510, 0.896)	0.842 (0.783, 0.898)	0.842 (0.783, 0.897)	0.045 (0.008, 0.088)	0.045 (0.007, 0.089)

Table 2-2: Parameters estimated by occupancy models applied to eDNA of ancient livestock (Giguet-Covex et al., 2014), ancient plants (Pansu, Giguet-Covex, et al., 2015), crayfish (Dougherty et al., 2016), salamanders (Vörös et al., 2017) and lampreys (Ostberg et al., 2018). S, sample size; K, number of PCRs; ψ_b, estimated baseline occupancy in the CAR-SODM; ψ_0, estimated occupancy probability in the Royle & Link (2006) model; p_{11}, estimated TP probability; p_{10}, estimated FP probability; α, estimated autoregressive parameter. CAR: CAR-SODM; RL: Royle & Link (2006) model. 95% Highest Probability Density Intervals (HPDIs) are presented in parentheses under each estimate.

Posterior conditional probabilities of occupancy were not identical between CAR-SODM and the Royle & Link (2006) models. For a given number of positive amplifications, CAR-SODM tended to give more support to the samples nearby to other positive detections, than to the ones in isolation (Figure 2-5), while the Royle & Link (2006) model assigned the same probability of occurrence to all samples sharing the same number of positive

amplifications. This discrepancy between the two models was most evident for samples with a single positive amplification. For example, in the lake sediment dataset, only 1/8 PCR detected *Bos* (i.e. cattle) eDNA within the sample at 343 cal. years BP. The CAR-SODM assigned to this sample a high (>0.9) posterior conditional probability of occupancy, as multiple positive detections occurred in the same period, while the Royle & Link (2006) model assigned a lower probability of occupancy (0.56) to the same sample (Figure 2-5c).

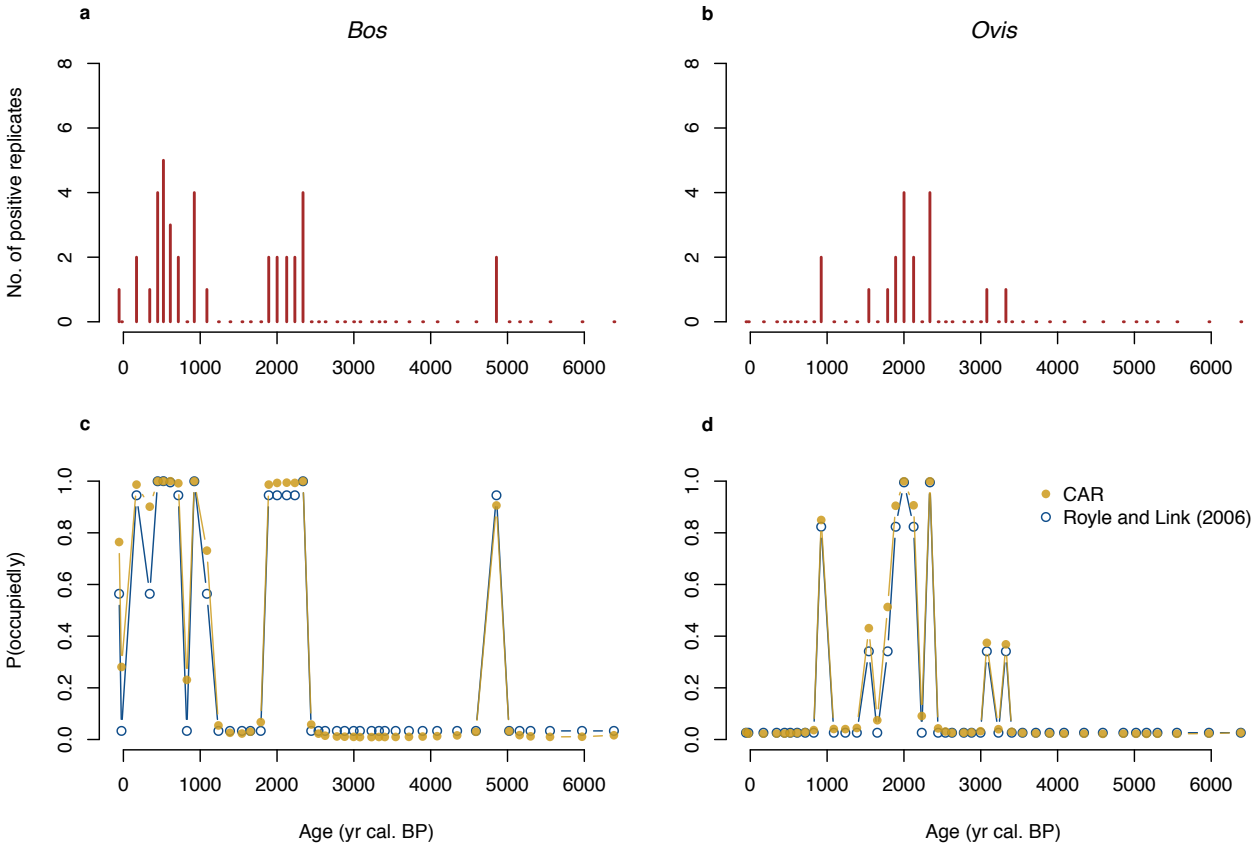


Figure 2-5: Applying SODMs to a temporally-structured ancient mammal DNA dataset (Giguet-Covex et al., 2014). a and b: Numbers of positive PCR replicates of *Bos* and *Ovis* eDNA, respectively. c and d: Corresponding estimated posterior conditional probabilities of occupancy, estimated by the CAR-SODM (full circles) and by the Royle & Link (2006) model (open circles).

2.4 Discussion

It is well known that both spatial and temporal autocorrelation can introduce biases to ecological data analyses (Beale et al., 2010; Brown et al., 2011; Lennon, 2000), and specifically, to species occupancy modeling (Johnson et al., 2013). Despite repeated calls (Ficetola et al., 2015; Lahoz-Monfort et al., 2016; Schmidt et al., 2013), only half of studies used SODM to analyze eDNA data, and few of them considered FPs in data analysis, though such issue might have been accounted for in their laboratory processing. Furthermore, until now very few studies using SODMs accounted for autocorrelation in spatially or temporally structured eDNA data, while the vast majority of papers did not report tests of autocorrelation before or after applying SODMs. Our study allows elucidating how biodiversity studies based on eDNA can benefit from SODMs taking both FPs and autocorrelation into account.

Detection errors are barely avoidable in biodiversity surveys, occur even with sessile species, and can be particularly problematic in environmental DNA studies (Guillera-Arroita, 2017). In cases where unambiguous detections are not available, the Royle & Link (2006) model is usually recommended to account for FPs (Ficetola et al., 2015, 2016; Lahoz-Monfort et al., 2016; Lopes et al., 2017). However, the fact that ecological datasets often show temporal or spatial autocorrelation requires the application of appropriate approaches. Here we show that a conditionally autoregressive model allows to successfully deal with autocorrelated occupancy data including both FPs and FNs, and that considering the autocorrelation among samples provides an improved statistical inference that can be extremely helpful in biodiversity studies.

In eDNA studies, the estimation of model parameters such as TP/FP probabilities is important not only for correctly predicting site occupancy but also for evaluating the appropriateness of molecular protocols (e.g. Lopes et al., 2017) and for measuring data reliability (Ficetola et al., 2015). In the occupancy models previously proposed for eDNA studies, the occupancy probability ψ_0 is one of such parameters (Guillera-Arroita et al.,

2017; Mackenzie et al., 2002; Miller et al., 2011; Royle & Link, 2006). However, if autocorrelation occurs this parameter is not constant for all samples. In the CAR-SODM, a fixed latent occupancy is replaced by the combination between a fixed baseline occupancy term ψ_0 and a varying autoregressive term (eqn. 3). As a result, the CAR-SODM was more flexible to fit autocorrelated data than its non-autoregressive counterpart and therefore less biased in estimating p_{11} and p_{10} (Figures 2-1 and 2-2). Both approaches tended to overestimate p_{10} in most scenarios, especially when K was small and p_{11} was low (Figure 2-2). The overestimation of p_{10} might be partially caused by the fixed uniform prior (0, 0.1), which may not be appropriate if the real p_{10} value is much smaller than the prior's upper limit 0.1, and if there is not enough information to determine p_{10} (e.g. when K is small). This would suggest that the p_{10} values might have been also overestimated for the empirical datasets (Table 2-2). However, this is probably not a major issue, given that in the empirical data all of the FPs are extremely low or have intervals that include zero. In real-world applications, one can adjust the prior based on knowledge about the FP rate in question, and information obtained from negative controls can help to provide realistic boundaries to p_{10}. A different pair of boundaries can be set to the uniform prior; alternatively, a beta prior can be applied instead (Gelman, Carlin, Stern, & Rubin, 2004), with parameters as the numbers of FP replicates and of true negative replicates in the control samples, when such information is available. Similarly, one can apply informative priors to other unknown parameters in the SODMs, potentially improving the accuracy and precision of their estimation. Additional analyses are required on the effects of prior specification for important model parameters, such as p_{11} and p_{10}. On the other hand, the baseline latent occupancy was poorly determined by the CAR-SODM (Table 2-2, note the large HPDIs), suggesting that the determination of this parameter demands a sample size much larger than here applied. This should not pose serious issues in most eDNA studies even with typical sample size (less than a hundred), if the estimation of this parameter is not the focus of the study. Conversely, if autocorrelation is weak (e.g. the salamander and lamprey datasets; Table 2-2), the CAR-SODM and the Royle & Link (2006) models provide highly consistent results. In this case, the Royle & Link (2006) model can have advantages, as its convergence is faster and it does not have the issue of the estimation of baseline latent occupancy.

It is no surprise that the CAR-SODM performs best when the samples are highly autocorrelated ($\alpha_c \geq 0.5$, Figs 3 and 4). Despite being high, these autocorrelation levels are frequent in real-world time datasets (e.g. Table 2-2), thus the application of CAR-SODM can be useful for many studies dealing with species occupancy data. CAR-SODM provided a particularly high improvement of performance when TP probability was low and only a few replicates were available for detection. Actually, these are the conditions under which inference is more challenging, and the use of SODM is essential to obtain unbiased biodiversity data. The performance of all the approaches clearly increased when more replicates are available, as shown by multiple studies on occupancy modeling (Ficetola et al., 2015; Lahoz-Monfort et al., 2016; Schmidt et al., 2013). For instance, with 12 replicates the maximum TSS of CAR-SODM can pass over 0.9 even if the TP rate was relatively low ($p_{11} = 0.25$) and the FP probability was as high as 0.015 (Figure 2-3). Given that scenarios of very low p_{11}, high p_{10} and autocorrelated sample are common in ancient DNA studies, the application of the CAR-SODM to such cases can be therefore promising. On the other hand, autoregressive models can be also helpful with moderate p_{11} (e.g. $p_{11} = 0.4$), particularly when only a few replicates are available ($K = 4$). The simulation study presented here also lends insights to the design of eDNA sampling and PCR schemes. For instance, when dealing with highly correlated chronosequence data ($\alpha_c = 0.95$) with low TP probability ($p_{11} = 0.25$), analyzing 8 PCR replicates with the autoregressive model ($K = 8$) provided nearly the same discriminating power provided by the non-autoregressive model with 12 PCR replicates ($K = 12$) (Figure 2-3), thus allowing to improve the performance of studies. On the other hand, it should be remarked that the same improvement could not be reached by the autoregressive model on data with just 4 replicates, compared with the non-autoregressive model on data with 8 replicates (Figure 2-3).

The CAR-SODM is an extremely flexible approach that can be applicable not only to time series data, but also to spatially explicit data. Actually, the performance of the model when dealing with spatially correlated data was coherent with the results of chronosequence analyses (compare Figures 2-2 and Figure 2-4a,b); see also Figure 2-3 and Figure 2-4c).

The only structural difference between these two applications is in the adjacency matrix **W**, that can be easily modified on the basis of researcher's expectations on how error is autocorrelated through space and time (Legendre & Gauthier, 2014). For instance, in our spatial example we used a rook scheme of adjacency, but different connection schemes can be used, for instance on the basis of known dispersal distance of study species.

Given that latent occupancy was not strictly constant, in CAR-SODM the probability of occupancy given a number of positive replicates can be variable, depending on whether the positive is nearby to other positives or not. For instance, one single detection of cattle eDNA is more likely to be considered a true presence if it occurs during periods in which cattle are frequently detected (Figure 2-5c). Considering the temporal coherence of observations has been frequently considered a good criterion to assess the validity of biodiversity data obtained through eDNA, particularly in ancient DNA studies (Giguet-Covex et al., 2014; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, Pedersen, et al., 2017). For instance, (Sjögren et al., 2017) considered positive PCR replicates of sedimentary plant DNA in two stratigraphically adjacent samples to be reliable, even if either or both of the two samples were amplified in just one PCR replicate. The outcome of CAR-SODMs is somehow analogous, given that they inherently take into account the temporal or spatial coherence of data, still CAR-SODMs provide a more objective approach to assess the status of samples for which target species have been detected only one or very few times. However, it should be remarked that the conditionally autoregressive model considered here assumed that autocorrelation is homogeneous through space and time, i.e. that the same autocorrelation values hold through the entire dataset. Such an assumption is common in spatially and temporally-explicit analyses, but autocorrelation can be non-homogeneous in real-world data (non-stationarity; Beale *et al.* 2010). Non-stationary of autocorrelation can greatly increase the bias of model outcomes, but unfortunately this issue remains challenging to address (Beale et al., 2010), and requires attention in future methodological developments.

The focus of the present work was on occupancy and detection probability estimation, therefore we compared the two-level Royle & Link (2006) model and its conditionally autoregressive counterpart, without considering any covariates to the model parameters. Nevertheless, the CAR-SODM can be easily modified, in order to meet specific needs. First, covariates can be incorporated to identify the drivers of temporal or spatial variations of TP/FP rates, and to take them into account. For example, DNA degradation may decrease detection probability of older samples (Olajos et al., 2017). Second, FPs and FNs can be generated by processes acting at different stages. For example, an FN may be caused by contamination during sampling, as well as at the PCR amplification stage. Therefore, multiple levels of latent states and corresponding parameters could be modeled instead of the singe level of the present work, especially when additional information is available (e.g. unambiguous detections or lab calibration data) (Guillera-Arroita et al., 2017).

2.5 Conclusions

Current occupancy modeling analyzing eDNA from spatially or temporally structured data ignore autocorrelation, despite many studies stressing the importance of autocorrelation in ecological data analysis. Using a conditionally autoregressive SODM, we showed when and how occupancy modeling in eDNA studies can benefit from considering the autocorrelation among samples. In comparison with the non-autoregressive Royle & Link (2006) model, the conditionally autoregressive model can better estimate important parameters such as TP/FP rates, and more accurately predict the actual occupancy of taxon, via discriminating detections according to their neighbors' states. The improvement was particularly high when the autocorrelation among samples was strong. We thus recommend using the autoregressive model in the frequent situations in which researchers expect autocorrelation among samples according to temporal or spatial structure, and when the TP rate is low.

2.6 Acknowledgements

We thank A. Kinziger and three anonymous reviewers for constructive comments on a previous version of this manuscript. Most of the computations presented in this paper were performed using the Froggy platform of the CIMENT infrastructure (https://ciment.ujf-grenoble.fr), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche.

GFF has received funding from the European Research Council under the European Community Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). The Laboratoire d'Écologie Alpine is part of Labex OSUG@2020 (ANR10 LABX56).

2.7 Data accessibility

R code for simulations, stan source code and commented examples of analyses of real-world datasets: https://gitlab.com/wtchen/DNA_CAR_model.git.

2.8 Author contributions

WC and GFF jointly designed this study. WC performed literature analysis, simulations and data analysis. WC wrote the first version of the manuscript, with subsequent contribution by GFF.

2.9 Reference

Aing, C., Halls, S., Oken, K., Dobrow, R., & Fieberg, J. (2011). A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated,

sites. Journal of Applied Ecology, 48(6), 1508–1517. http://doi.org/10.1111/j.1365-2664.2011.02037.x

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43(6), 1223–1232. http://doi.org/10.1111/j.1365-2664.2006.01214.x

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. Ecology Letters, 13(2), 246–264. http://doi.org/10.1111/j.1461-0248.2009.01422.x

Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Akre, T. S. B., Albaladejo, R. G., Albuquerque, F. S., … Hawkins, B. A. (2009). Coefficient shifts in geographical ecology: An empirical evaluation of spatial and non-spatial regression. Ecography, 32(2), 193–204. http://doi.org/10.1111/j.1600-0587.2009.05717.x

Bivand, R., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Guassian spatial autoregressive modles: An illustrated comparison of available methods. Geographical Analysis, 45(2), 150–179. http://doi.org/10.1111/gean.12008

Bivand, R., & Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software, 63(18). http://doi.org/10.18637/jss.v063.i18

Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145–1159. http://doi.org/10.1016/S0031-3203(96)00142-2

Brown, C. J., Schoeman, D. S., Sydeman, W. J., Brander, K., Buckley, L. B., Burrows, M., … Richardson, A. J. (2011). Quantitative approaches in climate change ecology. Global Change Biology, 17(12), 3697–3713. http://doi.org/10.1111/j.1365-2486.2011.02531.x

Capo, E., Debroas, D., Arnaud, F., & Domaizon, I. (2015). Is Planktonic Diversity Well Recorded in Sedimentary DNA? Toward the Reconstruction of Past Protistan Diversity. Microbial Ecology, 70(4), 865–875. http://doi.org/10.1007/s00248-015-0627-2

Chambert, T., Miller, D. A., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. Ecology, 96(2), 332–339. http://doi.org/10.1890/14-1507.1

Dorazio, R. M., & Erickson, R. A. (2018). ednaoccupancy: An r package for multiscale occupancy modelling of environmental DNA data. Molecular Ecology Resources, 18(2), 368–380. http://doi.org/10.1111/1755-0998.12735

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16(2), 129–138. http://doi.org/10.1111/j.1466-8238.2006.00279.x

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., … Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography, 30(5), 609–628. http://doi.org/10.1111/j.2007.0906-7590.05171.x

Dougherty, M. M., Larson, E. R., Renshaw, M. A., Gantz, C. A., Egan, S. P., Erickson, D. M., & Lodge, D. M. (2016). Environmental DNA (eDNA) detects the invasive rusty crayfish Orconectes rusticus at low abundances. Journal of Applied Ecology, 53(3), 722–732. http://doi.org/10.1111/1365-2664.12621

Epp, L. S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., … Brochmann, C. (2015). Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. Quaternary Science Reviews, 117(0318), 152–163. http://doi.org/10.1016/j.quascirev.2015.03.027

Ficetola, G. F., Manenti, R., De Bernardi, F., & Padoa-Schioppa, E. (2012). Can patterns of spatial autocorrelation reveal population processes? An analysis with the fire salamander. Ecography, 35(8), 693–703. http://doi.org/10.1111/j.1600-0587.2011.06483.x

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., … Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543–556. http://doi.org/10.1111/1755-0998.12338

Ficetola, G. F., Poulenard, J., Sabatier, D., Messager, E., Gielly, L., Leloup, A., … Arnaud, F. (2018). DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Science Advances. Science Advances. http://doi.org/10.1126/sciadv.aar4292

Ficetola, G. F., Romano, A., Salvidio, S., & Sindaco, R. (2018). Optimizing monitoring schemes to detect trends in abundance over broad scales. Animal Conservation, 21(3), 221–231. http://doi.org/10.1111/acv.12356

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources, 16(3), 604–607. http://doi.org/10.1111/1755-0998.12508

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian Data Analysis. Chapman Texts in Statistical Science Series. http://doi.org/10.1007/s13398-014-0173-7.2

Gelman, A., Rubin, D. B., Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7(4), 457–472. http://doi.org/10.1214/ss/1177013604

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., … Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications, 5, 3211. http://doi.org/10.1038/ncomms4211

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography, 40(2), 281–295. http://doi.org/10.1111/ecog.02445

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods in Ecology and Evolution, 8(9), 1081–1091. http://doi.org/10.1111/2041-210X.12743

Hines, J. E., Nichols, J. D., Royle, J. A., MacKenzie, D. I., Gopalaswamy, A. M., Samba Kumar, N., & Karanth, K. U. (2010). Tigers on trails: occupancy modelling for cluster sampling. Ecological Applications, 20(5), 1456–1466. http://doi.org/10.1890/09-0321.1

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. Biometrics, 61(4), 950–961. http://doi.org/10.1111/j.1541-0420.2005.00359.x

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., & Pond, B. A. (2013). Spatial occupancy models for large data sets. Ecology, 94(4), 801–808. http://doi.org/10.1890/12-0564.1

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. Molecular Ecology Resources, 16(3), 673–685. http://doi.org/10.1111/1755-0998.12486

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space − time analysis of community composition data. Proceedings of the Royal Society of London B: Biological Sciences, 281(1778), 20132728. http://doi.org/http://dx.doi.org/10.1098/rspb.2013.2728

Legendre, P., & Legendre, L. F. J. (2012). Numerical ecology (3rd ed.). Amsterdam: Elsevier.

Lennon, J. J. (2000). Red-shifts and red herrings in geographical ecology. Ecography, 23(1), 101–113. http://doi.org/10.1111/j.1600-0587.2000.tb00265.x

Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17(2), 145–151. http://doi.org/10.1111/j.1466-8238.2007.00358.x

Lopes, C. M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K. R., & Haddad, C. F. B. (2017). eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Molecular Ecology Resources, 17(5), 904–914. http://doi.org/10.1111/1755-0998.12643

Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew, J., & Langtimm, C. a. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. Ecology, 83(8), 2248–2255. http://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Miller, D. a., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. a. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. Ecology, 92(7), 1422–1428. http://doi.org/10.1890/10-1396.1

Olajos, F., Bokma, F., Bartels, P., Myrstener, E., Rydberg, J., Öhlund, G., … Englund, G. (2017). Estimating species colonization dates using DNA in lake sediment. Methods in Ecology and Evolution. http://doi.org/doi: 10.1111/2041-210X.12890

Ostberg, C. O., Chase, D. M., Hayes, M. C., & Duda, J. J. (2018). Distribution and seasonal differences in Pacific Lamprey and Lampetra spp eDNA across 18 Puget Sound watersheds. PeerJ, 6, e4496. http://doi.org/10.7717/peerj.4496

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. Molecular Ecology, 24(7), 1485–1498. http://doi.org/10.1111/mec.13136

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., … Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924–942. http://doi.org/10.1111/NPH.14470

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., … Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913–929. http://doi.org/10.1111/ecog.02881

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12(1), 77.

Royle, J. A., & Dorazio, R. (2009). Hierarchical Modeling and Inference in Ecology. Hierarchical Modeling and Inference in Ecology. http://doi.org/10.1016/B978-0-12-374097-7.X0001-4

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positives and false negative errors. Ecology, 87(4), 835–841. http://doi.org/10.1890/0012-9658(2006)87[835:GSOMAF]2.0.CO;2

Sargeant, G., Sovada, M., Slivinski, C., & Johnson, D. (2011). Markov Chain Monte Carlo estimation of species distributions: a case study of the swift fox in western Kansas. Journal of Wildlife Management, 69(2), 483–497. Retrieved from papers://8be6e07d-d85a-4bbf-8745-ec23972a8a55/Paper/p526

Schmidt, B. R., Kéry, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. Methods in Ecology and Evolution, 4(7), 646–653. http://doi.org/10.1111/2041-210X.12052

Serrao, N. R., Reid, S. M., & Wilson, C. C. (2017). Establishing detection thresholds for environmental DNA using receiver operator characteristic (ROC) curves. Conservation Genetics Resources, 1–8. http://doi.org/10.1007/s12686-017-0817-y

Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Merkel, M. K. F., … Alsos, I. G. (2017). Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytologist, 213(2), 929–941. http://doi.org/10.1111/nph.14199

Stan Development Team. (2016). {RStan}: the {R} interface to {Stan}. Retrieved from http://mc-stan.org/

Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems. Science, 240(4857), 1285–1293. http://doi.org/10.1126/science.3287615

Vörös, J., Márton, O., Schmidt, B. R., Gál, J. T., & Jelić, D. (2017). Surveying Europe's only cave-dwelling chordate species (proteus anguinus) using environmental DNA. PLoS ONE, 12(1), 12–14. http://doi.org/10.1371/journal.pone.0170945

Wagner, H. H., & Fortin, M. J. (2005). Spatial analysis of landscapes: Concepts and statistics. Ecology, 86(8), 1975–1987. http://doi.org/10.1890/04-0914

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference, 121(2), 311–324. http://doi.org/10.1016/S0378-3758(03)00111-3

2.10 Supporting Information

Supporting Information Table can be downloaded at https://doi.org/10.1111/ 1755-0998.12949

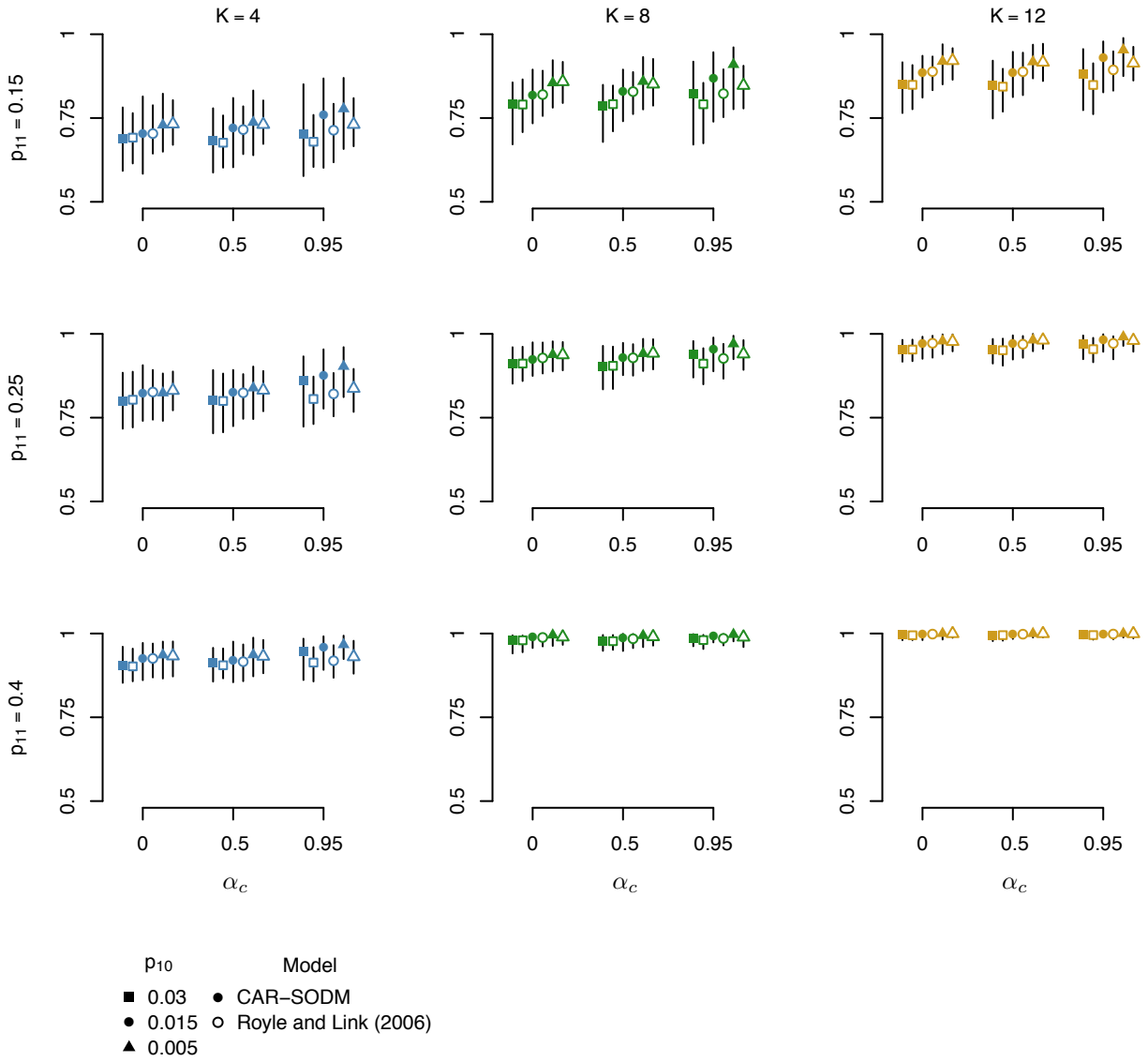


Figure 2-S1 Area under the receiver operating characteristic curve (AUC) for the CAR-SODM and the Royle and Link (2006) model, applied to the chronosequence simulated datasets. Results are arranged by scenarios of true positive (TP) probability p_{11}, autocorrelation level α_c and false positive (FP) probability p_{10}. Full symbols: CAR-SODM; open symbols: Royle and Link (2006). Plots show values calculated over 100 simulations. Vertical lines cover the 2.5%-97.5% quantiles.

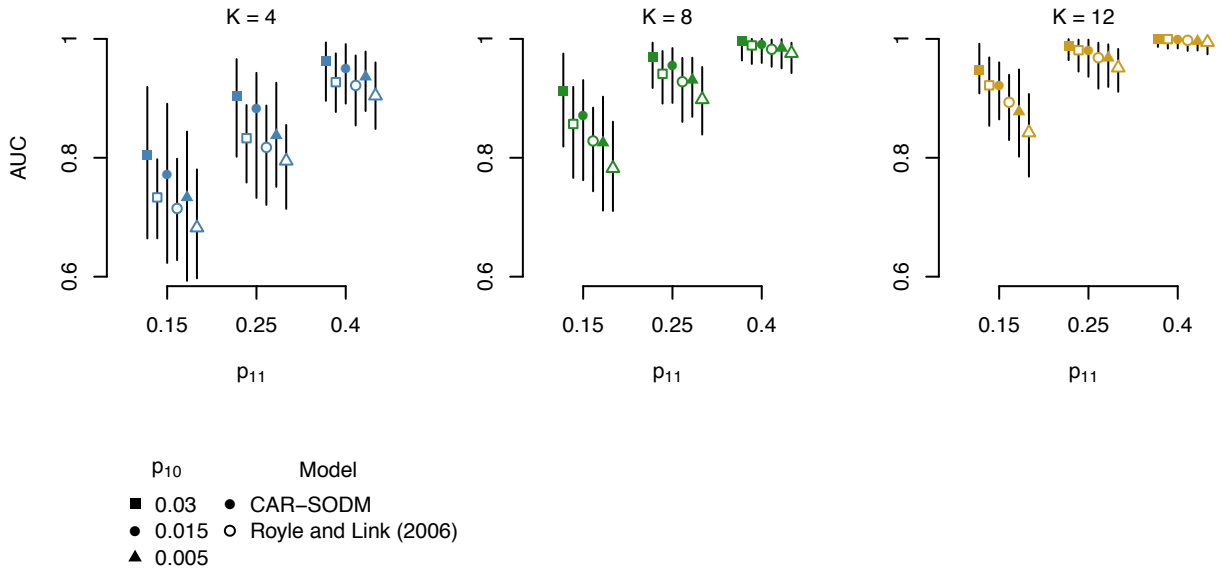


Figure 2-S2 Area under the receiver operating characteristic curve (AUC) for the CAR-SODM (solide symbols) and the Royle and Link (2006) model (open symbols). All results are arranged by scenarios of true positive probability p_{11} and false positive probability p_{10}.

Chapter 3. Lake sediment DNA reveals long-term impacts of livestock farming on plant community composition and variability (Article B)

Wentao Chen[1], Pierre Sabatier[2], Jérôme Poulenard[2], Fabien Arnaud[2], Erwan Messager[2], Pierre Taberlet[1], Manon Bajard[2], Delphine Rioux[1], Gentile Francesco Ficetola[1, 3]

[1] Univ. Grenoble Alpes, CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

[2] Environment Dynamics and Territories of Mountains (EDYTEM), University Savoie Mont Blanc, CNRS, 73000 Chambéry, France

[3] Department of Environmental Science and Policy, Università degli Studi di Milano. Via Celoria 26, 20133 Milano, Italy

Abstract

Paleoecological studies provide key information on how human activities and environmental disturbances affected past dynamics of vegetation in long time-scales, by reconstructing past human-environment interactions. Communities often respond to anthropogenic and environmental disturbances with long-term directional changes. Furthermore, theory predicts that disturbance can increase temporal variability of community composition, but to date tests of this hypothesis have been scarce, because of lack of exhaustive community data. Sedimentary environmental DNA (*Sed*DNA) is an emerging approach to jointly trace past vegetation changes and the presence of domestic mammals, thus providing accurate measure of the impact of human activities. Here we analyzed a *Sed*DNA record from a high-elevation lake in the Western Mediterranean Alps, France, to investigate the long-term dynamics of an alpine plant community and its relation to livestock farming and climate change over the last 6500 years, considering also how community variability changes through time. Neither livestock farming recorded in *Sed*DNA nor temperature change had strong directional impact on plant communities. However, variability of community composition was significantly higher during the Roman period (2076~1615 cal yr BP), which was the period with most frequent livestock farming. Vegetation dynamics in response to human disturbance and climate has a multi-dimensional nature, which determines not only directional changes in species composition, but can also affect the variability of community structure. Jointly considering directional and variability changes can provide essential insights on ecosystem shifts for a more complete understanding of past environmental changes.

Keywords: sediment environmental DNA, alpine vegetation, composition variability, biodiversity, livestock farming, climate change

3.1 Introduction

Alpine vegetation is highly sensitive to environmental changes of natural and anthropogenic origin (Körner, 2003). In the context of global climate change, high elevation flora is predicted to suffer strong impacts during the 21st century (Stocker, 2014) due to rise of temperature, change of precipitation pattern, longer growing season length and other related alterations (Dirnböck, Dullinger, & Grabherr, 2003; Engler et al., 2011), even if the impact of these factors on the distributions of particular species may be affected by microhabitat associations and microclimatic variability (Kulonen, Imboden, Rixen, Maier, & Wipf, 2018). In the Alps, recent observations have confirmed that such impact actually are ongoing [e.g. treeline shift (Leonelli et al., 2011), plant diversity change (Pauli et al., 2012) and plant species upward migration (Parolo & Rossi, 2008)]. On the other hand, the impacts of climate change on local plant diversity are known to be largely conditioned by microhabitat characteristics (Kulonen et al., 2018; Scherrer & Körner, 2011; Theurillat, 1995). Nevertheless, records on local vegetation change available through direct observation only cover a relatively short period of time (mostly over the last decades), so that the long-term trajectory of vegetation under climate is difficult to reveal. Furthermore, human activities have been modifying the landscapes in high altitude environments since prehistoric time. Indeed, human modifications of environment in the Alps are evident since at least 6600 years BP, as indicated by increased fire frequency (Tinner, Conedera, Ammann, & Lotter, 2005), with strong vegetation modifications even at high elevations (Pini et al., 2017).

The impacts of environmental changes on plant communities can be perceived through changes of vegetation composition, as already demonstrated by pollen records (Austrheim, Gunilla, Olsson, & Grontvedt, 1998; Bajard et al., 2015; Colombaroli, Beckmann, van der Knaap, Curdy, & Tinner, 2013). Most previous studies of past alpine vegetation focused on long-term diversity and composition shifts as a result of climate change and anthropogenic disturbances such as fire regime and livestock farming through millennia, e.g. (Colombaroli et al., 2013; Gehrig-Fasel, 2007; Tinner et al., 2005). The temporal compositional stability

(or its opposite, variability) of alpine communities under disturbance is rarely considered in paleoecological studies, even though similar topics have been largely investigated experimentally, e.g. (Klanderud & Totland, 2008; Speed, Austrheim, & Mysterud, 2013). In general, higher disturbance intensity is expected to lead to higher community variability (Fraterrigo & Rusak, 2008). For alpine grasslands, higher grazing disturbances often induce increased community compositional variability, but the actual response depends on factors such as moss cover (Speed, Cooper, Jónsdóttir, van der Wal, & Woodin, 2010), nutrient input (Hautier et al., 2014; Yang et al., 2012; Zhang et al., 2016), and invasion of shrub species (Báez & Collins, 2008). Especially, livestock farming is considered as an important disturbance, affecting community variability. For example, a 10-year field experiment performed in mountains in Southern Norway showed that in the presence of increased sheep grazing density, plant community temporal stability was increased at higher altitude but decreased at lower altitude (Speed, Austrheim, Hester, & Mysterud, 2013). Increased temporal variability often precedes regime shifts in an ecosystem (Carpenter & Brock, 2006; Kéfi, Dakos, Scheffer, Van Nes, & Rietkerk, 2013; Scheffer et al., 2009), but long-term information is needed to fully appreciate the relationships between disturbance, community variability and ecosystem shifts. High-resolution paleoecological data, recording how plant community variability changed in response to disturbances in the long-term, would be thus essential to understand ecosystem variation under disturbances. However, such paleoecological analyses on high-altitude environment currently are extremely scarce.

Given that multiple factors can impact high-altitude vegetation, a growing number of studies is trying to tease apart their relative role and importance. For instance, the effects of livestock farming, agriculture and even tourism on high altitude vegetation are sometimes assumed to be more severe and immediate than the ones of climate change (Körner, 2003), as their direct impacts are extremely strong. However, in the long term (e.g. through millennia), climate change has probably been an important driver. Some analyses on lakes at the treeline elevation suggest the human activities have been a more important driving factor than climate (Pansu, Giguet-Covex, et al., 2015; Pini et al., 2017). Even so, the situation above the treeline can be different. First, climate is a particularly important driver

of biodiversity limits at the upper limit of species' ranges (Whittaker, Willis, & Field, 2001). Second, at high altitudes, human impacts can be less strong (e.g. shorter growing season, shorter season available for human activities). In fact, herbivore exclusion experiments on alpine grasslands showed minor effect of grazing on community structure and underlined the climate influence on plant community response to grazing (Pardo, Doak, García-González, Gómez, & García, 2015). Unfortunately, there is very limited information on what happened in long term above the treeline.

To address these issues, it is possible to reconstruct past plant community dynamics using sedimentary archives. Palynological analysis, i.e. the identification and count of pollen and spores in sediment records, is the traditional approach to reveal the long-term vegetation history (Birks & Birks, 2000). Despite its power and popularity, the palynological approach is not always well adapted for studies focusing on a local scale. For many species, pollen grains are transported over long distances, therefore provide a regional idea of the vegetation composition, while potentially failing to infer local community composition (Boessenkool et al., 2014; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). This is probably the case for the reconstruction of past plant communities above-treeline, where the local signals could be overwhelmed by pollen from lower altitudes, due to the huge discrepancy in pollen production above and below tree lines (Birks & Birks, 2000). Besides, a lot of herb taxa with a low pollen production are rarely found in pollen records (Pardoe, 2001).

Recent advancement of high-throughput DNA sequencing technology allows an alternative approach to long-term vegetation studies, by uncovering the information stored in sediment environmental DNA (*Sed*DNA). *Sed*DNA can provide finer taxonomic resolution and capture higher richness than pollen (Niemeyer, Epp, Stoof-Leichsenring, Pestryakova, & Herzschuh, 2017), allowing the identification of species assemblage in the surrounding land ecosystem and correlate well with macrofossil records thanks to its mostly local origin (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). Recent studies also suggest that *Sed*DNA is an alternative to macrofossils, as they share similar origin and

largely overlap (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017), with *Sed*DNA being more sensitive than macrofossils in detecting less abundant taxa (Alsos, Sjögren, et al., 2016a). Therefore, *Sed*DNA is a useful tool to investigate past environmental changes complementary to traditional palynological approaches (Ficetola, Poulenard, et al., 2018; Jørgensen et al., 2012; Parducci et al., 2013; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017; Mikkel Winther Pedersen et al., 2013). *Sed*DNA mostly originates from species living within the lake watershed (Alsos, Lammers, Yoccoz, Jørgensen, et al., 2018). As a consequence, the results of *Sed*DNA and pollen analyses are often non-identical, and these should be considered as alternative tools that allow answering questions on processes acting at different spatial scales, i.e. local vs. regional (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). For instance, sedimentary DNA provides better information on the presence/absence of taxa in the vicinity of the lake (Alsos, Lammers, Yoccoz, Jørgensen, et al., 2018; Niemeyer et al., 2017).

Here we used *Sed*DNA to understand the long-term environmental evolution in an alpine area in the North-western Alps through the last 6500 years (Sabatier et al., 2017). First, we tested whether long-term environmental changes (climate, human activities) have determined vegetation changes, and assessed their relative importance through the millennia. Second, we tested whether grazing disturbances induced increased temporal variability of community composition.

3.2 Material and methods

3.2.1 Study site and coring

Lake Savine is located in the Haute-Maurienne massif of the French-Italian Alps (2447m a.s.l., Figure 3-1). The lake's catchment has a 3.5 km^2 surface, rising up to 3310 m. a.s.l. and is mainly (90%) made up by crystalline rock with Permian gneiss and micaschist. The

lake receives detrital inputs mainly from the southeastern part of the catchment from mid-June to mid-November, while the catchment is covered with snow and the lake is frozen the rest of the year. Geological and coring details are described in (Sabatier et al., 2017). Today, the lake's catchment area is covered by high altitude alpine meadows and siliceous screes. Nowadays, the meadows are occasionally grazed by cattle during summer.

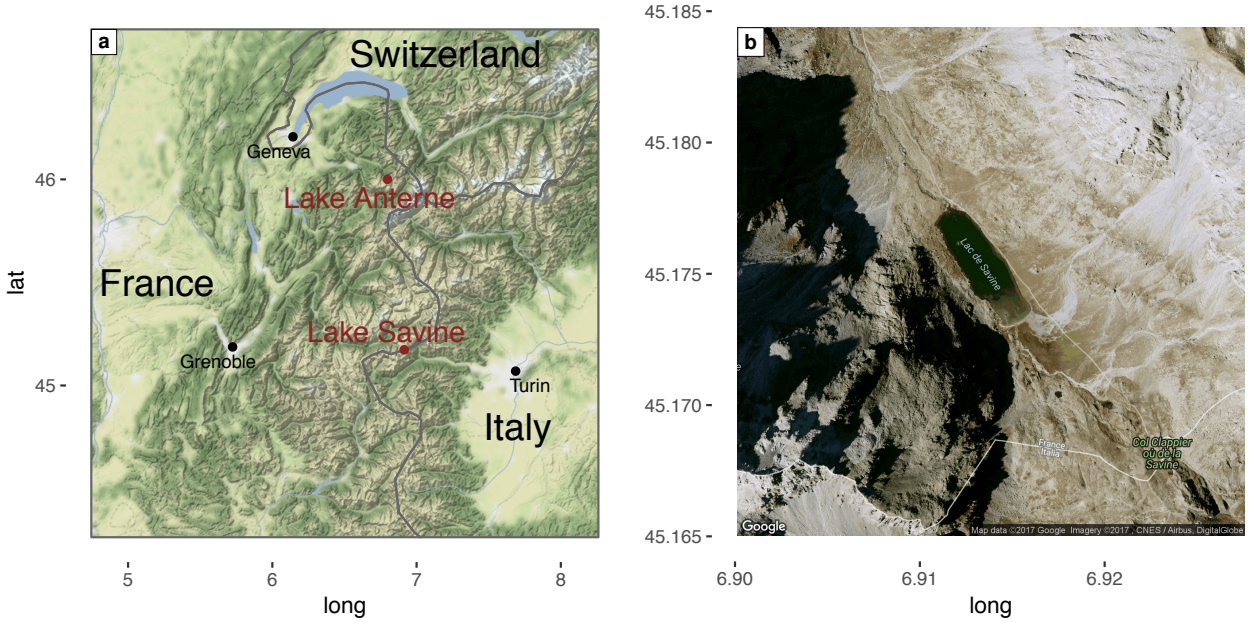


Figure 3-1: Location of Lake Savine in the Western Alps. Source: Google Maps, December 2017.

In March 2014, a 775-cm-long sediment core was extracted from the deepest part of Lake Savine (7m depth, N45°10.500, E6°54.821). Sabatier et al., (2017) presented a high-resolution sedimentological and geochemical analysis of the 775-cm-long sediment core sampled in Lake Savine, which led to the identification of 200 flood event layers and 20 underwater mass movements possibly related to earthquakes for the last 6500 years. The chronology of this sediment core results from 12 ^{14}C ages on terrestrial organic macro-remains and short-lived radionuclides (see Figure 3-S1 for the age model), for further detail see Sabatier et al., (2017).

3.2.2 Extraction, amplification and sequencing of extracellular DNA

From the sediment core, we sampled 72 slices (2-cm thickness) covering the whole core. Details of the sampling scheme (depths, thicknesses, weights) can be found in Table 3-S1. In order to avoid potential contamination, we removed the edges of slices with decontaminated, disposable scalpels. DNA extraction targeted extracellular DNA (Pansu, Giguet-Covex, et al., 2015). For each sediment slice, we mixed about 15 g of sediment with 15 ml of saturated phosphate buffer (Na2HPO4; 0.12 M; pH ≈ 8) during 15 min. 15 ml of the mixture was then centrifuged (10min at 10000 g). 12 ml of the resulting supernatant were transferred to Amicon ® Ultra-15 10K Centrifugal Filter Devices (Millipore) and centrifuged (20 min at 4000 g) for ultrafiltration and concentration of DNA. 400 µl of the resulting concentrate was kept as starting material for the following extraction steps, using the NucleoSpin® Soil kit (Macherey-Nagel, Düren, Germany), skipping the cell lysis step and following manufacturer's instructions [as described in (Taberlet, Prud'Homme, et al., 2012)].

Mammals and plants were the targets of DNA amplification. Mammal DNA was amplified with a primer pair targeting on a 60 to 84bp fragment of the mitochondrial 16S gene (P007 in Giguet-Covex et al., 2014; which corresponds to Mamm02 of Taberlet, Bonin, Zinger, & Coissac, 2018). For plant DNA, we used the primers targeting the P6 loop region of chloroplast *trn*L intron, a fragment whose length varies from 10 to 220 bp (*g*-*h* in Taberlet et al., 2007, which corresponds to Sper01 in Taberlet et al., 2018). To assign sequence reads to the relevant sample, 8 bp tags (with at least 5 differences between them) were added to the 5' end of primers (Binladen et al., 2007; Valentini, Pompanon, & Taberlet, 2009). DNA amplifications were carried out in a final volume of 20 µL containing 2 µL of diluted DNA extract. The amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer and 4.8 µg of bovine serum albumin (BSA, Roche Diagnostic). To limit the amplification of human DNA when using Mamm02 primers, we also added human-specific, blocking oligonucleotide (MamP007_B_Hum1) After 10 min at 95°C, we performed 45 PCR cycles: 30 s at 95°C, 30 s at 50°C, and 1 min

at 72°C with a final elongation step (7 min at 72°C) was performed. In addition to extraction controls, we also performed three PCR controls, containing PCR mix but no DNA template, and four PCR positive controls, each containing 0.18 ng of DNA extracted from a marsupial (*Didelphis marsupialis*, Didelphidae), absent in Europe. The positive control for plants was extracted from *Mouriri* sp., a tropical plant from the South America, belonging to a family (Melastomataceae) also absent in Europe. All samples and controls were amplified in 12 replicated PCRs (Ficetola et al., 2015). PCR products were then purified and mixed before sequencing. Sequencing was performed by 2 x 125 bp pair-end sequencing on Illumina HiSeq 2500 platform.

3.2.3 DNA sequence read filtering and taxonomic assignment

We filtered the DNA sequences adopting the protocol described in Pansu et al., (2015), using the OBITOOLS software package (Boyer et al., 2016). First, paired-end sequences were aligned and merged using the *illuminapairedend* program which implements a gap-free alignment algorithm. Identical sequences were merged together with their information of original samples being kept by using the *obiuniq* command. Merged sequences with alignment score lower than 40 were discarded. Each retained sequence was assigned to its original sample by the *ngsfilter* program. A sequence was retained only if it contained both primers with no more than two mismatches in each as well as the exact sample tag sequence. Using the *obigrep* command, we discarded all sequences shorter than 10 bp or containing ambiguous nucleotides. Sequences occurring only once in the dataset were discarded. To filter out potential amplification and sequencing errors, the *obiclean* program were used to create networks of sequences by linking any sequences differing by a single nucleotide substitution or indel. According to it corresponding network, each sequence fell into one of three categories: 'heads', which were only connected to sequences with a lower count, 'singletons', which were not connected to any other sequence and 'internals', which did not fall into either of the categories above. Sequences falling more often into 'internal' than 'head' or 'singleton' were then considered to be likely erroneous sequences and discarded.

Each remaining sequence formed a molecular operational taxonomic unit (MOTU). Taxonomic labels were assigned to each MOTU with the *ecotag* program by aligning them to a reference database. The database was generated by *in silico* PCR using the ecoPCR program (Ficetola et al., 2010) on the EMBL release 128 nucleotide database (*nt*). For plants, the taxonomic alignment was firstly performed on the global database including all taxa in the EMBL database and then on a database with only alpine species (Pansu, Giguet-Covex, et al., 2015). Inconsistences between the two resulting assigned datasets were resolved based on knowledge about the present local vegetation (Table 3-S3). We consider PCR replicates with more than 10 reads as positive detection, given that we consistently obtained much more plant than mammal reads. Only those MOTUs assigned to family or better taxonomic levels (genus, species etc.) with a perfect match (100%) and with more than 4000 reads and more than 10 positive PCR replicates over the whole dataset were retained. For mammals, we kept only the MOTUs with a match higher than 97% to a mammal genus and with more than 5 reads in a sample (Sabatier et al., 2017). MOTUs with more than three positive replicates in the control samples were removed from the dataset. MOTUs assigned to exotic taxa were obvious contaminants and also discarded (Table 3-S2). The remaining MOTUs were kept for data analyses.

3.2.4 Site occupancy modeling

In *Sed*DNA studies, both false absences (i.e. failure to detect a taxon when it is present) and false presences (i.e. detection of a taxon when it is actually absent, for instance because of contamination) are possible (Ficetola et al., 2015). Correctly inferring the presence of domestic animals is crucial in order to explore the relation between human activities and vegetation change. We adopted a site occupancy modeling approach to estimate the probability of a sediment sample being occupied, given the number of detections in each sample, using the Misclassification Model (Royle & Link, 2006):

$$z_i \sim Bernoulli(\psi),$$

$$y_i \sim binomial(z_i * p_{11} + (1 - z_i) * p_{10}, K),$$

where ψ is the occupancy probability for all samples; z_i the state of occupancy in sample i ($z_i = 1$ when the site is occupied, otherwise $z_i = 0$); y_i the number of positive detections in sample i; p_{11}, p_{10} the true and false detection probabilities, respectively. K denotes the number of observations and is regarded as number of PCRs in this study (Ficetola et al., 2015). The conditional probability that a detection history with detections comes from an occupied sample can be calculated by (Lahoz-Monfort et al., 2016):

$$P_i = P(z_i = 1 | y_i = y) = \frac{\psi_i p_{11}^y (1-p_{11})^{K-y}}{\psi_i p_{11}^y (1-p_{11})^{K-y} + (1-\psi_i) p_{10}^y (1-p_{10})^{K-y}}.$$

The model was coded in R and feeded to JAGS version 4.2.0 (Plummer, 2003) through the R2jags package (Su & Masanao Yajima, 2015) to obtain the posterior probability distributions of ψ, p_{11}, p_{10} and P_i using 3 MCMC chains of 10000 iterations for each MOTU. The Gelman-Rubin statistic for each parameter was computed to check chain convergence (Kéry, 2010).

3.2.5 Ecological analyses

We based our ecological analysis on a plant DNA dataset consisting of presence/absence values of the MOTUs remaining after the filtering step. The oldest samples from the core (up to *ca.* 8300 cal. years BP) were excluded because of lack of plant DNA. DNA detected from the remaining samples was transformed to presence/absence value and then combined for further analysis.

First, we used clustering to test whether the chronosequences of plant DNA formed different periods. We performed a stratigraphically constrained cluster analysis by the method of incremental sum of squares CONISS (Grimm, 1987) on the presence/absence Bray-Curtis dissimilarity structure. In preliminary analyses, the broken-stick model (Bennett, 1996a) established too many significant zones (>20), mostly due to very high variability found in samples dated to the Roman period (ca. 1600-2200 cal yr BP). We thus determined the number of major periods based on the clustering results, the presence of

domestic animal DNA and prior knowledge of local climate and human activities (see Table 3-1 for the criteria).

Notation	No. of samples	Date (cal. years BP	Defining criteria	Comments
A	17	-35~1429	Low variability in plant community composition (CONISS distances† , multivariate dispersions‡)	Middle age to present
B	13	1615~2076	Containing most animal DNA, high variability in plant community composition (CONISS distances†, multivariate dispersions‡)	Roman period
C	19	2155~3794	CONISS cluster†	Mediterranean climatic influences§; early human impacts?
D	10	4100~6509	CONISS cluster†	Atlantic climatic influences§

Table 3-1: Definitions of the four periods used to infer vegetation-environment relationships. † See Figure 3-2a; ‡ see Figure 3-4b; § following Sabatier et al., (2017).

Second, we assessed the relative importance of climate versus disturbance by domestic animals in determining community composition. To this end, we performed a redundancy

analysis (RDA) relating the presence/absence dataset of plant MOTUs to proxies of environmental changes. No local-scale reconstructions of temperature are available. Therefore, we used the reconstructed values of European continental temperature anomalies (Marsicek, Shuman, Bartlein, Shafer, & Brewer, 2018), which are a proxy of broad-scale (continental) variation of climate. The domestic animal *Sed*DNA (sum of positive PCR replicates of *Bos* and *Ovis*) was considered as an indicator of the presence of pastoralism.

Third, we tested whether disturbance increases the temporal variability of community composition. To do so, distance-based tests for homogeneity of multivariate dispersions described in (Anderson, 2006) were applied to the presence/absence dataset of plant DNA, contrasting different periods. In brief, a principal coordinates analysis (PCoA) was performed on the Bray-Curtis dissimilarity matrix. The original dissimilarity from each point to its group spatial median was recovered as the corresponding Euclidean distance in the principal coordinate space. To adjust for unequal sample size among periods, a sqrt(n/(n-1)) correction was applied to each distance in each periods (Stier, Geange, Hanson, & Bolker, 2013). Finally, *F*-tests by 9999 permutations of the least-absolute-deviation (LAD) residuals were performed to test the null hypothesis of equal dispersions from spatial medians among periods.

To confirm that our conclusions are not biased by extreme sedimentation events, we also explored the impacts of flood deposit proportion in sediment samples on plant DNA detection and consequently on the ecological analyses based on plant DNA. In brief, each sample's proportion of flood deposit was quantified (Sabatier et al., 2017). Linear regressions were performed to check the relationship between flood deposit percentage and plant DNA detection rate. We re-ran the RDA on the effects of temperature and of pastoralism on plant community composition, conditioned by flood deposit percentage. The impacts of flood deposit proportion on multivariate dispersion tests were explored by redoing the analysis on a reduced plant DNA dataset, i.e., the dataset with all samples having a flood deposit percentage over 70% removed. See appendix part II for additional details on these analyses.

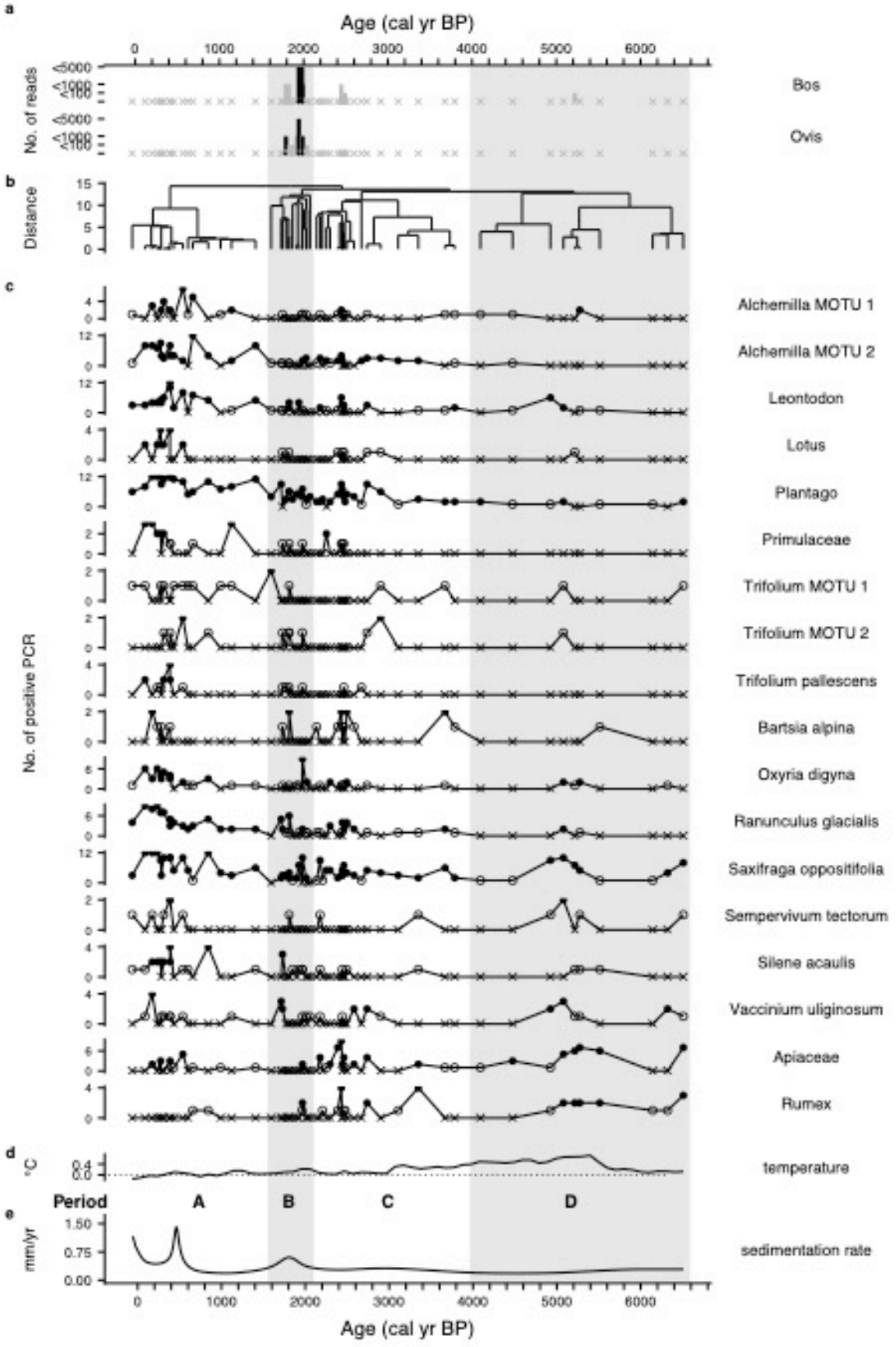

a
Age (cal yr BP)
0
1000
2000
3000
4000
5000
6000
No. of reads
Bos
Ovis
b
Distance
c
No. of positive PCR
Alchemilla MOTU 1
Alchemilla MOTU 2
Leontodon
Lotus
Plantago
Primulaceae
Trifolium MOTU 1
Trifolium MOTU 2
Trifolium pallescens
Bartsia alpina
Oxyria digyna
Ranunculus glacialis
Saxifraga oppositifolia
Sempervivum tectorum
Silene acaulis
Vaccinium uliginosum
Apiaceae
Rumex
d
°C
temperature
Period
A
B
C
D
e
mm/yr
sedimentation rate
Age (cal yr BP)

Figure 3-2: Development of indicator plant taxa over the last 6500 years with potential drivers of vegetation change. (**a**): domestic Animal (*Bos* and *Ovis*) DNA chronosequences; (**b**) CONISS clustering on plant assemblages; (**c**) development of vegetation with indicator taxa of alpine meadow (*Alchemilla* MOTU 1 and 2, *Leontodon*, *Lotus*, *Plantago*, Primulaceae, *Trifolium* MOTU 1 & 2 and *Trifolium pallescens*), scree (*Bartsia alpina*, *Oxyria digyna*, *Ranunculus glacialis*, *Saxifraga oppositifolia*, *Sempervivum tectorum* and *Silene acaulis*) and heathland (*Vaccinium uliginosum*); two plant taxa (Apiaceae and *Rumex*) that were abundant during the Holocene Climate Optimum but not in the more recent periods are also shown; (**d**) the reconstruction data of European continental temperature anomaly (Marsicek et al., 2018); (**e**) Sedimentation rate of the lake (Sabatier et al., 2017). Domestic animal DNA is represented by number of reads, detections with posterior occupancy probability >=0.5 are in black, otherwise in gray. Vegetation chronosequence is represented by the number of positive PCR replicates. Solid circle, open circle and cross represent >=2, one and no positive PCR replicates, respectively. Vegetation chronosequences, sedimentation rates and domestic animal DNA data are from the same sediment core (Sabatier et al., 2017).

All numerical analyses other than MCMC sampling were performed in the R environment version 3.3.2 (R Core Team, 2017). RDA and multivariate dispersion tests were performed by using the rda and betadisper functions, respectively, in the R package vegan version 2.4-2 (Oksanen et al., 2017). The chclust function in the R package rioja was used carry out the CONISS clustering (Juggins, 2015).

3.3 Results

3.3.1 SedDNA dataset

For the plant dataset, paired-end high-throughput sequencing yielded 35,306,021 reads over 924,120 unique sequences. 4,250,165 reads were found in the control samples (including positive controls). After filtering, 22,563,769 reads corresponding to 393 unique sequences remained in the dataset. Among these sequences, we retained only those found in at least 10 replicates, with at least 4,000 reads in total and taxonomically assigned to the family level or better. We discarded one sample dated at 5680 cal yr BP because it contained just 359 reads in two PCR replicates. Samples dedicated to test the impact of grain size on DNA contents were not included in the ecological analyses. In the final dataset, 10 MOTUs were assigned to the species level, 23 to the genus and 23 to the family/subfamily/tribe level (Table 3-S3).

The animal DNA dataset contained 40,638 reads distributed in 16 MOTUs after filtering. We found only two livestock species: *Bos* (8811 reads) and sheep *Ovis* (7152 reads). Most detections of these livestock MOTUs were concentrated in the samples dated from 1790 to 2040 cal. yr BP while single detections were found sporadically at around 1100 and 5200 cal. yr BP (Figure 3-2). Two closely dated samples at 2492 and 2458 cal yr BP, respectively, contained one out of twelve PCR replicates cattle DNA.

3.3.2 Temporal changes of communities

Our final dataset included 61 samples, which covered the period from -35 to 6509 cal. BP. Stratigraphically-constrained cluster analysis identified four major periods through this time-scale (Table 3-1). Period A comprised the 17 most recent samples up to 1429 cal yr BP, which showed relatively homogeneous community compositions (Figure 3-2, a and c). Period B included 14 samples dated from 1615 to 2076 cal yr BP. All detections of DNA of domestic animals with posterior probability of occupancy > 0.5 were found in this period (Figure 3-2b). The division of Periods C (20 samples, 2155~3794 cal yr BP) and D (10 samples, 4100~6509 cal yr BP) coincides with the abrupt shift of local hydroclimate from Atlantic to Mediterranean influences at 4kyr cal BP, suggested by (Sabatier et al., 2017) based on flood frequency reconstructed from the same sediment core used in the present study.

Through the whole chronosequence (Figs 2 and S3), the landscape was dominated by herbs Asteraceae MOTU 1 Poeae MOTUs. Period D was characterized by herbaceous taxa Apiaceae and *Rumex*, accompanied by some indicator taxa of alpine/rocky habitats, such as purple saxifrage *Saxifraga oppositifolia* and common houseleek *Sempervivum tectorum*. Period C was characterized by the rise of *Plantago* and *Alchemilla* MOTU 2 and the decline of some previously dominating taxa, e.g. *Saxifraga* MOTUs and Apiaceae, suggesting rising influences of human activities. Under constant disturbance of domestic animals, the vegetation in Period B was characterized by very high taxonomic turnover rate as the plant assemblage oscillated between two states. One state was characterized by

dominating *Plantago*, while other taxa were barely present. The second state was similar to the assemblages in late Period C and Period A, characterized by a more species-rich alpine meadow community (Figure 3-S3). The vegetation in Period A was dominated by meadow/grassland taxa, such as *Plantago*, *Alchemilla* MOTUs and Primulaceae, as well as some taxa of alpine and nival habitats, e.g. *Ranunculus glacialis*, *Saxifraga oppositifolia* and mountain sorrel *Oxyria digyna*. Bog blueberry *Vaccinium uliginosum* were frequently, but not continuously, present through all the periods, indicating the presence of some shrubs within heathland/moorland habitats.

3.3.3 Environmental predictors of community changes

Redundancy analysis (RDA) on the chronosequences allowed identifying important factors related to community changes (Figure 3-3). Community changes through time were significantly related to the reconstruction of regional variation of temperature (P =0.0053), which explained 3.73% of the total variance. Lower temperature were associated cold-adapted taxa such as *Ranunculus glacialis* (Figure 3-3b). However, we detected only a few warm-adapted taxa,with warmer temperature, namely Apiaceae, *Euphrasia* and *Rumex*. Conversely, RDA did not detect significant relationship between community composition and pastoralism (1.27% of the total variance, P=0.7814). The RDA axes together explained only 5% of the total variance. Temporal autocorrelations of residuals was significant (at α = 0.05) for just 4/56 species (7.1%), a proportion similar to the expected type-1 error rate, suggesting that temporal autocorrelation was not an issue to our analyses.

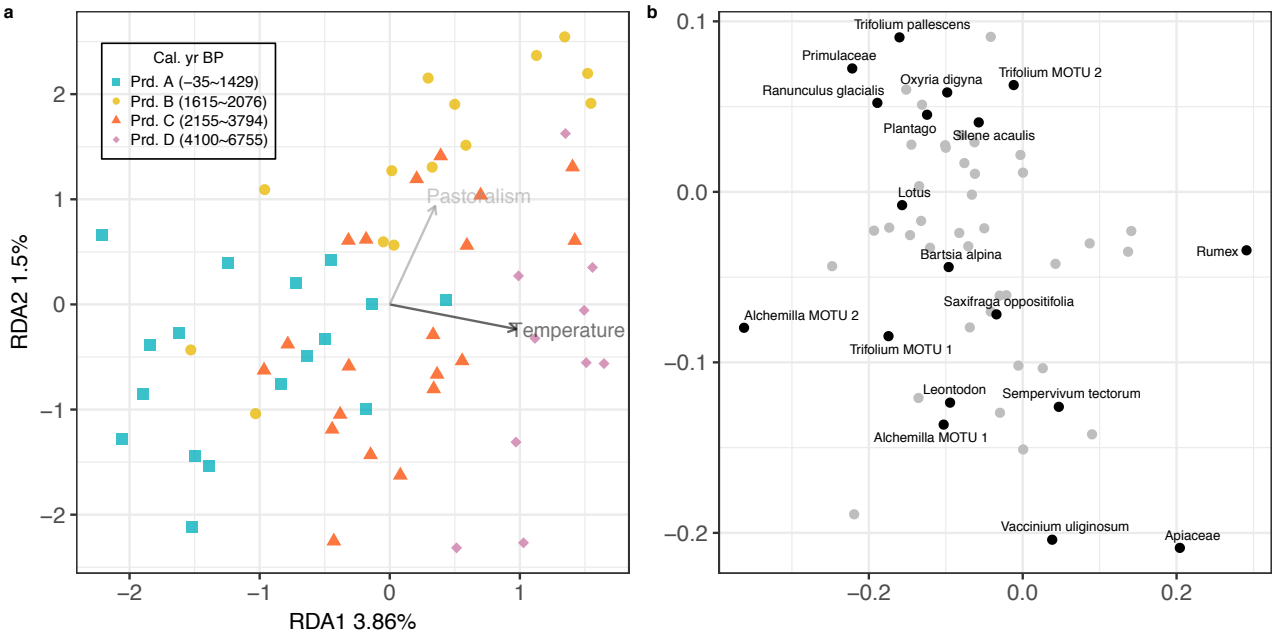


Figure 3-3: Redundancy analysis (RDA) on past plant assemblages. **(a)** RDA axes of the plant assemblages (5.36% of the total inertia). Arrows show explanatory variables. **(b)** Plant taxa ordination on RDA axes. Plant taxa presented in Figure 3-2 are named in black. Other taxa are only represented by plots in gray for clarity.

3.3.4 Disturbance and temporal variability of communities

The dispersions of plant assemblages within each period are shown in a plot of the first two principal coordinates from the dissimilarity matrix (Figure 3-4). Plant assemblages belonging to the Period B showed significantly higher variance compared to communities of both Periods A (Permutation tests: P=0.0002) and C (Permutation tests: P=0.0165). The smallest average distance to the spatial median of the corresponding period was found in Period A, while the largest one in the period B (Figure 3-4b). In particular, the average distance to the spatial median was significantly larger for assemblages of Period B than that of Period A and to a less degree, than that of Period C (Table 3-2).

Period pair	$F_{\text{permutation}}$	$P_{permutation}$

A-B	-4.80	0.0002
A-C	-2.97	0.0047
A-D	-3.38	0.0010
B-C	2.57	0.0165
B-D	1.65	0.1151
C-D	-0.54	0.5999

Table 3-2: Pairwise comparisons for multivariate dispersions among plant assemblages, using binary Bray-Curtis dissimilarities from spatial medians. No adjustment for multiple comparisons was applied. F_m values were obtained from 9999 permutations of LAD residuals.

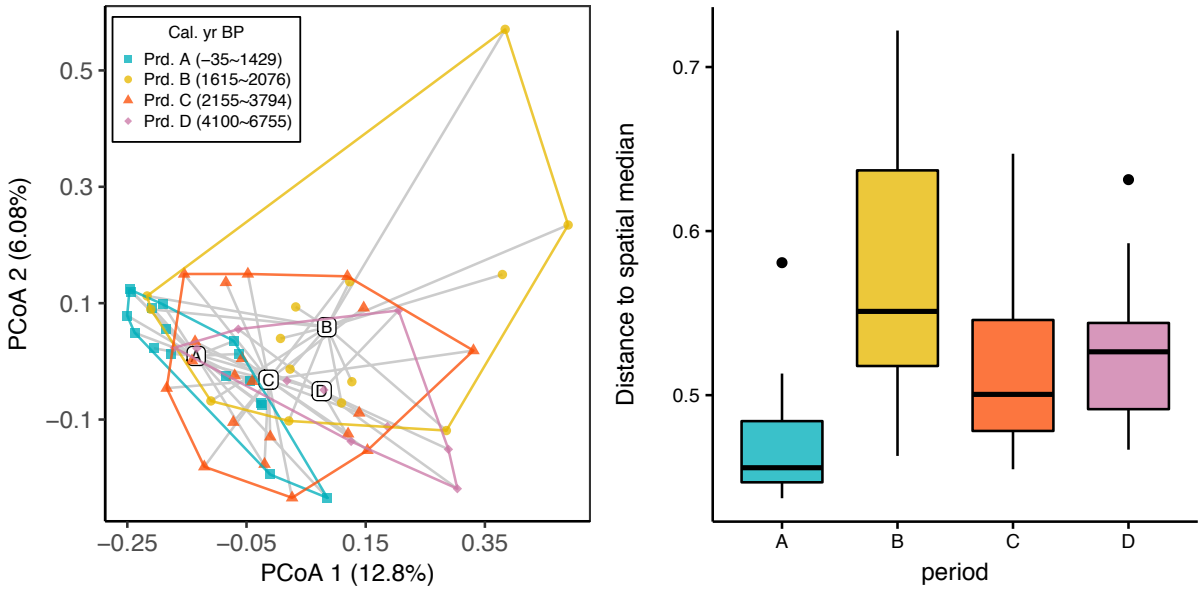


Figure 3-4: Multivariate dispersions based on the binary Bray-Curtis dissimilarities between sediment samples. **(a)** First two principal coordinate axes on the dissimilarity matrix. Samples belonging to period 1-4 are represented by squares, circles, triangles and diamonds, respectively. Numbers denote projections of spatial medians of corresponding periods. **(b)** Boxplots of the samples' distances to the spatial median of the corresponding period in the principal coordinate space.

Tests on the relationship of flood deposit proportion in sediment samples with plant DNA detection revealed a mild effect, and the occurrence of flood deposits did not significantly

affect the major patterns discovered by RDA and multivariate dispersion tests (see Supplementary Information, Tables 3-S5~S7, Figures 3-S6~S7 for details).

3.4 Discussion

3.4.1 Vegetation dynamics in high altitudes

*Sed*DNA has been proven to be able to provide fine-scale paleoecological information, thanks to its local origin (Ficetola, Poulenard, et al., 2018; Pansu, Giguet-Covex, et al., 2015; Parducci et al., 2013; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). In this study, we used *Sed*DNA to reconstruct the plant community evolution of an alpine grassland in the French-Italian Alps through the last 6500 years. The altitude of the study area may probably have been above the forest line, even during the Holocene Climate Optimum (Carcaillet, 1998), which can explain why we did not find evidence of pristine forest communities before the first human-induced disturbances. As a result, vegetation changes under the impact of human activities were less drastic, compared to the plant community transition from forest to grazed land through a similar time scale at lower altitude, reconstructed for other Alpine lakes (Giguet-Covex et al., 2014; Pansu, Giguet-Covex, et al., 2015).

Major transition of vegetation occurred at the beginning of Period C (~4000 cal yr BP) with the rise of taxa characterizing dense alpine meadows, e.g. *Plantago*, *Alchemilla* MOTUs and *Leontodon*, which is accompanied by the decline of some previously dominating taxa, e.g. Apiaceae and *Rumex* (Figure 3-2c). This trend intensified after around 2600 cal yr BP, in the period corresponding with the earliest records of the Celto-Ligurian civilization in this region (Chastagnol, 1970). The first stage of this transition may be therefore the earliest evidence of the impact of human land use (slash and burn cultivation and low-intensity pastoralism) in the Lake Savine area. However, the detection of *Sed*DNA of domestic animals was only sporadic in this period, and conspicuous changes did not appear until Period B (i.e. 2076 to 1615 cal yr BP). This corresponds to the Roman period, when DNA

sequences of domestic animals were continuously found (Figure 3-2a). Plant assemblage composition reflected by sediment DNA experienced increased temporal variability during this period, as it oscillated between a species-poor and a species-rich state (Figures 3-4 and 3-S3, also Figure 3-2b). The transition arrived at its final stage at the beginning of Period A (1429 cal yr BP), when plant assemblages became dominated by taxa that indicate densely vegetated meadows (e.g. *Alchemilla* MOTUs and *Leontodon*), with some alpine taxa such as *Ranunculus glacialis* and *Saxifraga oppositifolia* representing some rocky habitats around the lake (Figure 3-2c). The plant assemblages were stabilized at such state during the whole Period A, as its compositional variability remained minimum (Figures 3-2b and 3-4). This alpine meadow landscape of Period A conforms to those observed in surrounding areas at present.

3.4.2 Role of climate change

Alpine plant communities are supposed to be highly sensitive to climate changes (Körner, 2003). The present climate pattern in the study area is similar to the European continental climate (Carcaillet, 1998; Fournier, 1985), with influences from the Mediterranean sea during mesoscale high precipitation events (Durand et al., 2009). Given the lack of local-scale climatic reconstruction, our analysis only included a proxy of broad-scale regional temperature variation (European continental temperature anomalies, Marsicek et al., 2018), which represents well the overall temperature change occurring across Europe over the Holocene. However, as its continental scale suggest, this curve may not always reflect the climatic variations in an alpine area in the Western Alps. Besides, its 100-year interval may fail to capture higher-frequency temperature variation during the periods where time intervals between adjacent sediment samples were much shorter than 100 years (e.g. during period B and a large part of period C). Indeed, the available temperature data only explained a small fraction of the variance (<5%) observed in the Lake Savine plant assemblages. This result was unlikely affected by the flood deposit contents in the samples (see Appendix part II, Table 3-S6). Besides, only a few taxa that dominated the plant assemblages during the Holocene climate optimum positively responded to temperature

increase, while most other taxa, regardless of their thermal preferences, were negatively correlated to temperature (Figure 3-3b). Such distribution suggests that the apparent effect of temperature detected by RDA might even be a result of other covariates not analyzed in this study. Apart from the limits of the temperature proxy, a possible explanation to the relatively small impact of temperature on vegetation change is that microclimatic characteristics and microhabitat conditions may be a more important factor determining plant alpine species distribution (Kulonen et al., 2018). Indeed, topographical heterogeneity can often defy regional temperature trend in determining plant species distribution in alpine environment, by differentiating thermal condition on nearby sites (Choler, 2017; Scherrer & Körner, 2010, 2011).

Finally, the strong heteroscedasticity (heterogeneity of variability) across the periods (Figure 3-4b) may have affected the RDA results, which is a statistical method assuming homogeneity of variance (Legendre & Legendre, 2012). Temporal variation in community composition evolution data reconstructed from environmental DNA can occur as result of the DNA degradation process through time. Future methodological developments on *Sed*DNA data analysis should address this issue.

3.4.3 Impacts of livestock farming

Although the oldest high-altitude seasonal camps of pastoral use in the Alps can be dated as far as to the Neolithic (Curdy, 2007), our *Sed*DNA results could not reliably establish the stable presence of livestock in the Lake Savine area earlier than 2500 cal. years BP (Figure 3-2). Similarly, despite the sporadic presence of cattle around the lake during the summer at present day, no DNA read of domestic animals was found in recent samples. This may be due to the fact that only relatively intensive livestock farming can provide a sufficient amount of mammal DNA in sediment that can be detected (Giguet-Covex et al., 2014). The detection of cattle DNA in two closely dated samples at around 2500 cal yr BP may indicate the onset of pastoral activities, but the highest amount of cattle and sheep DNA reads was found in samples from Period B (Figure 3-2a). This indicates a peak of livestock

farming that is consistent with the intensive production of cheese in the Centronian Alps (situated in the Savoie region of modern France) during the Roman period, reported by several authors in antiquity (Chastagnol, 1970; Giguet-Covex et al., 2014). We therefore hypothesize that intensive livestock farming occurred during the Roman period. As the plant DNA data indicates the presence of communities of grazed alpine meadow through Period A, we cannot exclude the possibility that some pastoral activities persisted from the beginning of the Middle Ages up to present day.

RDA is among the most efficient approaches to detect the factors affecting community changes, even within a temporal framework, but it assumes a linear relationship between species and environmental factors (Legendre & Gauthier, 2014; Legendre & Legendre, 2012). However, our results suggest that the impact of livestock on vegetation was more complex, with an increase in species turnover rate rather than a directional change in composition. Despite RDA did not detect strong and direct effects of livestock farming on plant community composition (Figure 3-3a), vegetation dynamics during the Roman period (Period B) was highly perturbed by cattle and sheep grazing, as attested by the eDNA of both mammals (Figs 2c and 4). The higher community variability during Period B can be observed by the longer branches in the CONISS cluster tree (Figure 3-2) and also by the higher degree of plant assemblage dispersion (Figure 3-4). The increased variability during the Period B is not caused by increased sedimentation or variability of sedimentary conditions (see Figures 3-S4~S6), as these parameters were not more extreme during the Roman Period than during Period A (Figure 3-2e). Furthermore, the effects of flood deposit content on the test results of multivariate dispersion were small and did not affect the detected patterns (see Appendix, Table 3-S7), and the increased variability during Period B is unlikely determined by age-related DNA degradation in the sediment samples, because the variability in Period B was higher than both the periods before (C) and after (A).

Overall, we found a much higher variability of plant communities along the Period B, compared to the other periods, and this is likely the result of intensive pastoral activities. Modern practice of transhumance in the local area observes a series of rules to avoid

deteriorating the alpine meadows. For example, in some pastures with a very long annual period of snow cover (8-9 months), grazing period must not begin until August, to allow the complete installation of the vegetation that ensure the sustainability of the meadow (Bornard, Bassignana, Bernard-Brunet, Labonne, & Cozic, 2006). Conversely, Roman pastoral practice, if lacking effective guidelines to protect the meadows, could have repeatedly driven the local plant communities into the species-poor state observed here, followed by rebounds to species-rich states if the deterioration of pasture determined a reduction of farming intensity. Indeed, anthropogenic disturbances (including livestock grazing) have been considered as a major cause of increased variability in ecosystems (Hautier et al., 2015). More intensive disturbances, especially those of long durations, may induce higher community variability via reducing species abundances and then leaving newly available resources to poorer competitors (Fraterrigo & Rusak, 2008). Increased spatial heterogeneity created by grazing may have also contributed to the temporal variability surge (Collins et al., 2018), because it is likely that pastoral activities during the Period B were limited to the flat grassland part of the catchment (Sabatier et al., 2017).

Increased variability is one of the early warning signals for critical transitions in many complex systems (Carpenter & Brock, 2006; Scheffer et al., 2009; Spanbauer et al., 2014). It is therefore an extensively discussed topic concerning ecosystem resilience and regime shift. The present paper reports a record of variability-increasing event of alpine vegetation dynamics that could have been a leading indicator of regime shift. Our study was unable to identify other early warning signals such as longer return time of state variables after a perturbation (Scheffer et al., 2009, 2012), perhaps because of limited temporal resolution, or because the ecosystem never showed a drastic shift to typical pastureland communities. Nevertheless, we highlight the capability of *Sed*DNA data to reconstruct not only directional changes in ecological patterns, but also variability shifts, which allow researcher to gain valuable insights to the complex interactions among human activities, climate and ecosystems. These capabilities can also provide excellent opportunities to test ecological theories (e.g. alternative stable state theory (Scheffer et al., 2009)) on a large time scale that is seldom possible to reach by experiments nor by historical records.

3.5 Acknowledgements

We thank Dr. Charline Giguet-Covex for her information on the present pastoralism status in the Lake Savine area. This research was performed by the Hannibal Project funded by the DIPEE (Dispositifs de Partenariat en Ecologie et Environnement) Chambéry Grenoble of the CNRS INEE. We thank the municipality of Bramans for the coring authorization. The authors thank CLIMCOR Continent for the coring facilities. 14C analyses were acquired thanks to the CNRS-INSU ARTEMIS national radiocarbon AMS measurement programme at Laboratoire de Mesure 14C (LMC14) in the CEA Institute at Saclay (French Atomic Energy Commission). GFF has received funding from the European Research Council under the European Community Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). We thank Philippe Gaucher for froviding tissue of *Didelphis marsupialis*. The Laboratoire d'Écologie Alpine is part of Labex OSUG@2020 (ANR10 LABX56).

3.6 Data Accessibility

- Raw and filtered *sed*DNA (mammal and plant) sequencing data, plant reference database and geochemical and sedimentological data are available in the Dryad Digital Repository
- The reconstructed values of European continental temperature anomalies are available in the NOAA National Centers for Environmental Information Paleoclimatology Database: https://www1.ncdc.noaa.gov/pub/data/paleo/reconstructions/marsicek2018/TemperatureReconstructions2.csv.

3.7 Author Contributions

W.C., G.F.F., F.A. and P.T. conceived the study. P.S. F.A. and J.P. performed the fieldwork. P.S performed the sedimentological analyses. G.F.F., P.T. and D.R. performed the laboratory work on *sed*DNA. W.C. performed the analyses with subsequent contributions by G.F.F., E.M., P.S., J.P. and M.B. W.C. wrote the first draft of the manuscript, with subsequent contribution by all the authors.

3.8 Conflict of interests

P.T. is one of the co-inventors of patents related to the Sper01 (*g*-*h*) primers and the use of the P6 loop of the chloroplast *trn*L (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers.

3.9 References

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. Plos One, 13(4), e0195403. http://doi.org/10.1371/journal.pone.0195403

Alsos, I. G., Sjögren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., … Pedersen, M. W. (2016). Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene, 26(4), 627–642. http://doi.org/10.1177/0959683615612563

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62(1), 245–253. http://doi.org/10.1111/j.1541-0420.2005.00440.x

Austrheim, G., Gunilla, E., Olsson, A., & Grontvedt, E. (1998). Land-use impact on plant communities in semi-natural sub-alpine grasslands of Budalen, central Norway. Biological Conservation, 87(3), 369–379. http://doi.org/10.1016/S0006-3207(98)00071-8

Báez, S., & Collins, S. L. (2008). Shrub invasion decreases diversity and alters community stability in Northern Chihuahuan desert plant communities. PLoS ONE, 3(6). http://doi.org/10.1371/journal.pone.0002332

Bajard, M., Sabatier, P., David, F., Develle, A. L., Reyss, J. L., Fanget, B., … Arnaud, F. (2015). Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. Holocene, 26(3), 350–364. http://doi.org/10.1177/0959683615609750

Bennett, K. D. (1996). Active and passive environmental DNA surveillance of aquatic invasive species 2 3. New Phytologist, 132(1), 155–170. http://doi.org/10.1111/j.1469-8137.1996.tb04521.x

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PloS One, 2(2), e197.

Birks, H. H., & Birks, H. J. B. (2000). Future uses of pollen analysis must include plant macrofossils. Journal of Biogeography, 27(1), 31–35. http://doi.org/10.1046/j.1365-2699.2000.00375.x

Boessenkool, S., McGlynn, G., Epp, L. S., Taylor, D., Pimentel, M., Gizaw, A., … Popp, M. (2014). Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. Conservation Biology, 28(2), 446–455. http://doi.org/10.1111/cobi.12195

Bornard, A., Bassignana, M., Bernard-Brunet, C., Labonne, S., & Cozic, P. (2006). Les végétations d'alpage de la Vanoise Description agro-écologique et gestion pastorale. (J. Baudel, Ed.) (Collection). Editions Quae.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. http://doi.org/10.1111/1755-0998.12428

Carcaillet, C. (1998). A spatially precise study of Holocene fire history, climate and human impact within the Maurienne valley, North French Alps. Journal of Ecology, 86, 384–396. http://doi.org/10.1046/j.1365-2745.1998.00267.x

Carpenter, S. R., & Brock, W. A. (2006). Rising variance: A leading indicator of ecological transition. Ecology Letters, 9(3), 308–315. http://doi.org/10.1111/j.1461-0248.2005.00877.x

Chastagnol, A. (1970). La province romaine des Alpes cottiennes. Annales. Histoire, Sciences Sociales, 5, 1318–1320. Retrieved from http://www.jstor.org/stable/27577732

Choler, P. (2017). Winter soil temperature dependence of alpine plant distribution: Implications for anticipating vegetation changes under a warming climate. Perspectives in Plant Ecology, Evolution and Systematics, (October), 0–1. http://doi.org/10.1016/j.ppees.2017.11.002

Collins, S. L., Avolio, M. L., Gries, C., Hallett, L. M., Koerner, S. E., La Pierre, K. J., … Jones, M. B. (2018). Temporal heterogeneity increases with spatial heterogeneity in ecological communities. Ecology, 99(4), 858–865. http://doi.org/10.1002/ecy.2154

Colombaroli, D., Beckmann, M., van der Knaap, W. O., Curdy, P., & Tinner, W. (2013). Changes in biodiversity and vegetation composition in the central Swiss Alps during the transition from pristine forest to first farming. Diversity and Distributions, 19(2), 157–170. http://doi.org/10.1111/j.1472-4642.2012.00930.x

Curdy, P. (2007). Prehistoric settlement in middle and high altitudes in the Upper Rhone Valley (Valais-Vaud, Switzerland): A summary of twenty years of research. Preistoria Alpina, 42(December), 99–108.

Dirnböck, T., Dullinger, S., & Grabherr, G. (2003). A regional impact assessment of climate and land use change on alpine vegetation. Journal of Biogeography, 30, 1–17. http://doi.org/10.1046/j.1365-2699.2003.00839.x

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., & Mérindol, L. (2009). Reanalysis of 44 yr of climate in the French Alps (1958-2002): Methodology, model validation, climatology, and trends for air temperature and precipitation. Journal of Applied Meteorology and Climatology, 48(3), 429–449. http://doi.org/10.1175/2008JAMC1808.1

Engler, R., Randin, C. F., Thuiller, W., Dullinger, S., Zimmermann, N. E., Araújo, M. B., … others. (2011). 21st century climate change threatens mountain flora unequally across Europe. Global Change Biology, 17(7), 2330–2341. http://doi.org/10.1111/j.1365-2486.2010.02393.x

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 1. http://doi.org/10.1186/1471-2164-11-434

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., … Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543–556. http://doi.org/10.1111/1755-0998.12338

Ficetola, G. F., Poulenard, J., Sabatier, D., Messager, E., Gielly, L., Leloup, A., … Arnaud, F. (2018). DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Science Advances. Science Advances. http://doi.org/10.1126/sciadv.aar4292

Fournier, J. (1985). Contribution à l'étude des Alpes intermédiaires françaises: la Moyenne-Maurienne: bioclimatologie, groupements végétaux forestiers et impacts humains.

Fraterrigo, J. M., & Rusak, J. A. (2008). Disturbance-driven changes in the variability of ecological patterns and processes. Ecology Letters, 11(7), 756–770. http://doi.org/10.1111/j.1461-0248.2008.01191.x

Gehrig-Fasel, J. (2007). Tree line shifts in the Swiss Alps: Climate change or land abandonment? Journal of Vegetation Science, 18(4), 571–582. http://doi.org/10.1658/1100-9233(2007)18[571:TLSITS]2.0.CO;2

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., … Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications, 5, 3211. http://doi.org/10.1038/ncomms4211

Grimm, E. C. (1987). Constrained Cluster Analysis By the Method of Incremental Sum of Squares. Computers and Geosciences, 13(1), 13–35. http://doi.org/10.1016/0098-3004(87)90022-7

Hautier, Y., Seabloom, E. W., Borer, E. T., Adler, P. B., Stanley Harpole, W., -Lincoln Hautier, N., … Hector, A. (2014). Eutrophication weakens stabilizing effects of diversity in natural grasslands. Nature, 508(7497), 521–525. http://doi.org/https://doi.org/10.1038/nature13014

Hautier, Y., Tilman, D., Isbell, F., Seabloom, E. W., Borer, E. T., & Reich, P. B. (2015). Anthropogenic environmental changes affect ecosystem stability via biodiversity. Science, 348(6232), 336–340. http://doi.org/DOI: 10.1126/science.aaa1788

Jørgensen, T., Haile, J., Möller, P. E. R., Andreev, A., Boessenkool, S., Rasmussen, M., … Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Molecular Ecology, 21(8), 1989–2003. http://doi.org/10.1111/j.1365-294X.2011.05287.x

Juggins, S. (2015). rioja: Analysis of Quaternary Science Data. Retrieved from http://www.staff.ncl.ac.uk/stephen.juggins/

Kéfi, S., Dakos, V., Scheffer, M., Van Nes, E. H., & Rietkerk, M. (2013). Early warning signals also precede non-catastrophic transitions. Oikos, 122(5), 641–648. http://doi.org/10.1111/j.1600-0706.2012.20838.x

Kéry, M. (2010). Introduction to WinBUGS for ecologists. Academic, Burlington.

Klanderud, K., & Totland, Ø. (2008). Diversity-Stability Relationships of an Alpine Plant Community under Simulated Environmental Change. Arctic, Antarctic, and Alpine Research, 40(4), 679–684. http://doi.org/10.1657/1523-0430(07-075)

Körner, C. (2003). Alpine plant life: functional plant ecology of high mountain ecosystems; with 47 tables (2nd ed.). Springer Science & Business Media.

Kulonen, A., Imboden, R. A., Rixen, C., Maier, S. B., & Wipf, S. (2018). Enough space in a warmer world? Microhabitat diversity and small-scale distribution of alpine plants on mountain summits. Diversity and Distributions, 24(2), 252–261. http://doi.org/10.1111/ddi.12673

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. Molecular Ecology Resources, 16(3), 673–685. http://doi.org/10.1111/1755-0998.12486

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space − time analysis of community composition data. Proceedings of the Royal Society of London B: Biological Sciences, 281(1778), 20132728. http://doi.org/http://dx.doi.org/10.1098/rspb.2013.2728

Legendre, P., & Legendre, L. F. J. (2012). Numerical ecology (3rd ed.). Amsterdam: Elsevier.

Leonelli, G., Pelflni, M., Cella, U. M. Di, Garavaglia, V., Pelfini, M., Cella, U. M. Di, & Garavaglia, V. (2011). Climate warming and the recent treeline shift in the European Alps: the role of geomorphological factors in high-altitude sites. AMBIO: A Journal of the Human Environment, 40(3), 264–273. http://doi.org/10.1007/s13280-010-0096-2

Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., & Brewer, S. (2018). Reconciling divergent trends and millennial variations in Holocene temperatures. Nature, 554(7690), 92–96. http://doi.org/10.1038/nature25464

Niemeyer, B., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzschuh, U. (2017). A comparison of sedimentary DNA and pollen from lake sediments in recording vegetation composition at the Siberian treeline. Molecular Ecology Resources, 17(6), e46–e62. http://doi.org/10.1111/1755-0998.12689

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. Molecular Ecology, 24(7), 1485–1498. http://doi.org/10.1111/mec.13136

Pardo, I., Doak, D. F., García-González, R., Gómez, D., & García, M. B. (2015). Long-term response of plant communities to herbivore exclusion at high elevation grasslands. Biodiversity and Conservation, 24(12), 3033–3047. http://doi.org/10.1007/s10531-015-0996-3

Pardoe, H. S. (2001). The representation of taxa in surface pollen spectra on alpine and sub-alpine glacier forelands in southern Norway. Review of Palaeobotany and Palynology, 117(1–3), 63–78. http://doi.org/10.1016/S0034-6667(01)00077-X

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist.

Parducci, L., Matetovici, I., Fontana, S. L., Bennett, K. D., Suyama, Y., Haile, J., … Willerslev, E. (2013). Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Molecular Ecology, 22(13), 3511–3524. http://doi.org/10.1111/mec.12298

Parolo, G., & Rossi, G. (2008). Upward migration of vascular plants following a climate warming trend in the Alps. Basic and Applied Ecology, 9(2), 100–107. http://doi.org/10.1016/j.baae.2007.01.005

Pauli, H., Gottfried, M., Dullinger, S., Abdaladze, O., Akhalkatsi, M., Alonso, J. L. B., … Grabherr, G. (2012). Recent Plant Diversity Changes on Europe's Mountain Summits. Science, 336(6079), 353–355. http://doi.org/DOI 10.1126/science.1219033

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., … Kjær, K. H. (2013). A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quaternary Science Reviews, 75, 161–168. http://doi.org/10.1016/j.quascirev.2013.06.006

Pini, R., Ravazzi, C., Raiteri, L., Guerreschi, A., Castellano, L., & Comolli, R. (2017). From pristine forests to high-altitude pastures: an ecological approach to prehistoric human impact on vegetation and landscapes in the western Italian Alps. Journal of Ecology. http://doi.org/10.1111/1365-2745.12767

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 20–22. http://doi.org/10.1.1.13.3406

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positives and false negative errors. Ecology, 87(4), 835–841. http://doi.org/10.1890/0012-9658(2006)87[835:GSOMAF]2.0.CO;2

Sabatier, P., Wilhelm, B., Ficetola, G. F., Moiroux, F., Poulenard, J., Develle, A.-L., … Arnaud, F. (2017). 6-kyr record of flood frequency and intensity in the western Mediterranean Alps -- Interplay of solar and temperature forcing. Quaternary Science Reviews, 170, 121–135. http://doi.org/10.1016/j.quascirev.2017.06.019

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., … Sugihara, G. (2009). Early-warning signals for critical transitions. Nature, 461(7260), 53–59. http://doi.org/10.1038/nature08227

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., … Vandermeer, J. (2012). Anticipating critical transitions. Science, 338(6105), 344–348. http://doi.org/10.1126/science.1225244

Scherrer, D., & Körner, C. (2010). Infra-red thermometry of alpine landscapes challenges climatic warming projections. Global Change Biology, 16(9), 2602–2613. http://doi.org/10.1111/j.1365-2486.2009.02122.x

Scherrer, D., & Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. Journal of Biogeography, 38(2), 406–416. http://doi.org/10.1111/j.1365-2699.2010.02407.x

Spanbauer, T. L., Allen, C. R., Angeler, D. G., Eason, T., Fritz, S. C., Garmestani, A. S., … Stone, J. R. (2014). Prolonged instability prior to a regime shift. PLoS ONE, 9(10), 3–9. http://doi.org/10.1371/journal.pone.0108936

Speed, J. D. M., Austrheim, G., Hester, A. J., & Mysterud, A. (2013). The Response of Alpine Salix Shrubs to Long-Term Browsing Varies with Elevation and Herbivore Density. Arctic, Antarctic, and Alpine Research, 45(4), 584–593. http://doi.org/10.1657/1938-4246-45.4.584

Speed, J. D. M., Austrheim, G., & Mysterud, A. (2013). The response of plant diversity to grazing varies along an elevational gradient. Journal of Ecology, 101(5), 1225–1236. http://doi.org/10.1111/1365-2745.12133

Speed, J. D. M., Cooper, E. J., Jónsdóttir, I. S., van der Wal, R., & Woodin, S. J. (2010). Plant community properties predict vegetation resilience to herbivore disturbance in the Arctic. Journal of Ecology, 98(5), 1002–1013. http://doi.org/10.1111/j.1365-2745.2010.01685.x

Stier, A. C., Geange, S. W., Hanson, K. M., & Bolker, B. M. (2013). Predator density and timing of arrival affect reef fish community assembly. Ecology, 94(5), 1057–1068. http://doi.org/10.1890/11-1983.1

Stocker, T. (2014). Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Su, Y.-S., & Masanao Yajima. (2015). R2jags: Using R to Run "JAGS." Retrieved from https://cran.r-project.org/package=R2jags

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14--e14. http://doi.org/10.1093/nar/gkl938

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., … Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology, 21(8), 1816–1820. http://doi.org/10.1111/j.1365-294X.2011.05317.x

Theurillat, J.-P. (1995). Climate change and the alpine flora: some perspectives. In Potential Ecological Impacts of Climate Change in the Alps and Fennoscandian Mountains (pp. 121--127).

Tinner, W., Conedera, M., Ammann, B., & Lotter, A. F. A. F. (2005). Fire ecology north and south of the Alps since the last ice age. The Holocene, 15(8), 1214–1226. http://doi.org/10.1191/0959683605hl892rp

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. Trends in Ecology & Evolution, 24(2), 110–117. http://doi.org/10.1016/j.tree.2008.09.011

Whittaker, R. J., Willis, K. J., & Field, R. (2001). Scale and species richness : towards a general, hierarchical theory of species diversity. Journal of Biogeography, 28, 453–470.

Yang, H., Jiang, L., Li, L., Li, A., Wu, M., & Wan, S. (2012). Diversity-dependent stability under mowing and nutrient addition: Evidence from a 7-year grassland experiment. Ecology Letters, 15(6), 619–626. http://doi.org/10.1111/j.1461-0248.2012.01778.x

Zhang, Y., Loreau, M., Lü, X., He, N., Zhang, G., & Han, X. (2016). Nitrogen enrichment weakens ecosystem stability through decreased species asynchrony and population stability in a temperate grassland. Global Change Biology, 22(4), 1445–1455. http://doi.org/10.1111/gcb.13140

3.10 Supplementary Information

3.10.1 Supplementary tables and figures

Table 3-S1. Details of the sediment DNA samples. Flood- and earthquake deposit percentages were determined by eye inspecting and measuring the respective thicknesses in each sediment sample.

Sample	Best age	Age sup	Age inf	No. of PCR functional replicates	Depth top (cm)	Depth bottom (cm)	MCD (cm)	Weight (g)	Sample thickness (cm)	Organic matter %	Carbonate %	Detrital %	Event in the sample %	Event type	Comment
S01	-35	-39	-31	12	1	4	4	11.35	3	4.20	6.64	89.16	23	flood	
S02	117	83	148	12	13	16.5	16.5	4.19	3.5	2.52	11.13	86.35	71	flood	
S03	202	133	246	12	24.5	27.5	27.5	13.54	3	3.15	7.75	89.10	60	earthquake	

S04	262	16	31	12									100	flood
		6	3		39	42	42	6.79	3	2.84	8.26	88.90		
S05	302	18	35	12									100	flood
		7	9		62	64.5	56	19.25	2.5	3.15	7.78	89.06		
S06	310	19	36	12									100	flood
		2	9		76	78.5	70	11.36	2.5	4.26	7.09	88.65		
S07	339	21	40	11									100	flood
		1	0		87	89.5	81	11.55	2.5	4.52	6.90	88.58		
S08	414	26	48	12									100	flood
		2	3		97.5	100	91.5	16.68	2.5	3.76	6.63	89.61		
S09	414	26	48	12									100	flood
		2	3		116	118.5	110	18.51	2.5	3.09	8.18	88.73		
S10	458	31	53	12									0	
		1	8		128.5	131	122.5	11.87	2.5	4.20	7.73	88.08		
S11	566	47	66	12									20	flood
		8	4		141	143.5	135	12.29	2.5	5.27	9.62	85.11		
S12	627	55	71	12									33	flood
		7	9		15.5	18.5	143	8.41	3	3.89	7.31	88.80		
S13	684	62	77	12									28	flood
		5	0		27.5	30	154.5	10.96	2.5	4.38	7.29	88.33		
S14	867	81	92	12									100	flood
		4	6		39.5	42	166.5	11.86	2.5	4.15	7.89	87.97		
S15	101	96	10	12									0	
	4	1	64		51.5	54	178.5	11.21	2.5	7.32	6.80	85.88		

S16	1096	1037	1146	12	87	89.5	214	13.09	2.5	4.62	7.95	87.43	100	flood	fine grain flood deposit. not analyzed
S17	1096	1037	1146	12	100	103	227.5	14.35	3	2.61	7.88	89.51	83	flood	coarse grain flood deposit. not analyzed
S18	1144	1083	1196	12	122	125	249.5	17.98	3	4.24	7.91	87.86	67	flood	
S19	1429	1362	1487	12	132.5	135	259.5	13.63	2.5	4.74	8.68	86.59	100	flood	
S20	1615	1544	1668	12	74	76.5	271	18.83	2.5	4.92	7.21	87.87	8	flood	
S21	1734	1663	1790	12	87.5	90	284.5	17.5	2.5	5.40	6.84	87.77	32	flood	
S22	1749	1680	1807	11	90	92.5	287	15.05	2.5	4.76	6.92	88.33	64	flood	
S23	1790	1725	1855	12	92.5	95	289.5	13.35	2.5	5.61	7.59	86.80	0		
S24	1825	1764	1900	11	95	97.5	292	17.91	2.5	5.85	6.46	87.70	20	flood	
S25	1828	1768	1905	12	97.5	100	294.5	15.42	2.5	3.05	8.02	88.93	92	flood	
S26	186	18	19	12	100	102.5	297	15.53	2.5	5.87	7.25	86.88	0		

	9	13	62											
S27	191	18	20	11									0	
	7	68	22		102.5	105	299.5	12.44	2.5	5.37	7.13	87.50		
S28	194	18	20	12									48	flood
	3	95	54		105	107.5	302	14.3	2.5	4.21	6.99	88.80		
S29	198	19	20	12									36	flood
	1	32	97		107.5	110	304.5	16.72	2.5	4.90	6.28	88.82		
S30	198	19	21	12									88	flood
	6	37	03		110	112.5	307	14.24	2.5	3.72	7.04	89.24		
S31	202	19	21	11									32	flood
	9	74	47		112.5	115	309.5	15.67	2.5	4.85	7.95	87.19		
S32	204	19	21	11									84	earthq
	0	84	57		115	117.5	312	15.67	2.5	6.44	6.72	86.84		uake
S33	207	20	21	12									48	flood
	6	14	88		117.5	120	314.5	13.12	2.5	7.07	6.44	86.49		
S34	215	20	22	12									0	
	5	75	60		120	122.5	317	13.35	2.5	7.22	7.49	85.29		
S35	219	21	22	12									60	flood
	5	09	99		122.5	124.5	319	11.32	2	6.00	7.13	86.87		
S36	222	21	23	11									60	flood
	6	33	25		124.5	127	321.5	15.72	2.5	5.56	7.75	86.68		
S37	227	21	23	12									44	flood
	1	68	70		127	129.5	324	14.54	2.5	4.89	8.33	86.79		
S38	231	21	24	12	129.5	132	326.5	15.08	2.5	4.62	8.06	87.32	48	flood

	4	99	17												
S39	240	22	25	12									0		
	5	71	13		132	134.5	329	13.6	2.5	6.23	7.67	86.10			
S40	244	23	25	12									48	flood	
	8	08	61		134.5	137	331.5	12.1	2.5	6.24	8.19	85.57			
S41	245	23	25	12									56	earthq uake	
	9	16	72		137	139.5	334	14.84	2.5	6.23	6.55	87.22			
S42	245	23	25	12									100	earthq uake	
	9	16	72		139.5	142	336.5	13.35	2.5	5.99	7.76	86.25			
S43	247	23	25	12									72	flood	
	7	30	92		142	144.5	339	10.03	2.5	4.93	6.81	88.26			
S44	249	23	26	12									80	flood	
	2	41	09		144.5	147	341.5	11.33	2.5	5.23	7.38	87.39			
S45	249	23	26	12									100	flood	fine grain flood deposit. not analyzed
	2	41	09		147	149	343.5	10.64	2	2.49	8.19	89.32			
S46	249	23	26	12									100	flood	coarse grain flood deposit. not analyzed
	2	41	09		149	151.5	346	12.83	2.5	3.73	12.28	83.99			
S47	251	23	26	12									72	flood	
	3	57	32		151.5	154	348.5	9.89	2.5	7.56	9.12	83.33			
S48	259	24	27	12									0		
	8	27	23		158.5	161	355.5	11.18	2.5	6.79	7.09	86.12			

S49	268	25	28	12									60	flood
	7	09	21		6	8	369.5	8.93	2	8.57	9.02	82.41		
S50	275	25	28	12									28	earthq
	3	81	93		23.5	26	387.5	11.31	2.5	5.56	7.82	86.62		uake
S51	291	27	30	12									24	flood
	5	48	57		35.5	38	399.5	9.54	2.5	5.64	8.00	86.36		
S52	312	29	32	12									0	
	3	77	57		47.5	50	411.5	9.53	2.5	7.01	7.11	85.88		
S53	336	32	34	12									0	
	2	41	93		67	70	431.5	11.09	3	6.16	6.92	86.92		
S54	367	35	38	12									24	flood
	9	73	06		78.5	81	442.5	11.38	2.5	5.77	6.94	87.29		
S55	379	36	39	12									40	flood
	4	89	18		92.5	95	456.5	13.47	2.5	5.31	6.42	88.27		
S56	410	40	42	12									0	
	0	06	16		104.5	107	468.5	12.68	2.5	7.49	7.07	85.44		
S57	448	43	45	12									8	flood
	2	70	94		52.5	55	481	12.7	2.5	6.44	7.31	86.25		
S58	493	47	50	12									40	flood
	1	83	47		73.8	76.3	502.3	11.99	2.5	5.04	7.03	87.93		
S59	508	49	52	12									64	flood
	5	27	02		14	16.5	514.5	15.11	2.5	4.00	6.93	89.06		
S60	522	50	53	12									100	earthq
	0	62	35		25	27.5	525.5	13.2	2.5	3.14	6.52	90.33		uake

S61	5279	5121	5397	12	36.5	39	537	11.4	2.5	4.49	7.22	88.29	100	flood	
S62	5517	5367	5636	12	49.5	52	550	11.4	2.5	7.46	7.10	85.44	0		
S63	5680	5546	5801	12	60	62.5	560.5	13.5	2.5	6.10	6.30	87.60	16	flood	too few DNA reads. not analyzed
S64	6145	6023	6298	11	74	76.5	574.5	11.8	2.5	5.24	6.50	88.26	0		
S65	6322	6144	6529	12	85	87.5	585.5	11.9	2.5	4.45	6.08	89.47	56	earthquake	
S66	6509	6276	6779	12	97	99.5	597.5	12.8	2.5	3.36	6.98	89.66	52	flood	
S67	6755			12	113.5	116	614	11.88	2.5	4.67	6.56	88.77	56	flood	Unreliable date. not analyzed
S68	7022			12	125.5	128	626	13.1	2.5	4.35	7.21	88.44	28	flood	Unreliable date. not analyzed
S69	7363			12	136	138.5	636.5	11.5	2.5	4.82	7.26	87.91	0		Unreliable date. not analyzed
S70	7755			12	5.5	8	648.5	10.3	2.5	4.66	7.01	88.33	32	flood	Unreliable date. not analyzed

S71	8116	12	18.5	21	661.5	15.4	2.5	3.73	7.46	88.80	60	flood	Unreliable date. not analyzed
S72	8365	12	31	33.5	674	16.8	2.5	5.75	7.84	86.41	0		Unreliable date. not analyzed
S73		11	78	80.5	721	14.9	2.5	1.90	6.77	91.32			Unreliable date. not analyzed
S74		12	121	123.5	764	13.2	2.5	2.09	6.88	91.03			Unreliable date. not analyzed

Table 3-S2. Contaminant (more reads found in control samples than in sediment samples or >3 positive PCR replicates) and exotic plant DNA sequences.

	Scientific name (alpine database)	Scientific name (global database)	sequence	Sequence length	No. of reads in control samples	No. of detections in control samples
1	Poaceae	-	atccctttttgaaaaaacaagtggttct caaactagaacccaaatgaaaag	52	125951	1
2	Moraceae	-	atccggttttctgaaaacaaacaaggg ttcagaaagcgataatacaaaag	50	35202	1
3	-	Canavalia ensiformis	atcctattttccgaagacaaagaaaag ttcagaaagcgagaatataaaaag	51	48426	2
4	Saliceae	-	atcctatttttcgaaaacaaacaaaggt tcataaagacagaataagaatacaaa ag	56	625	12
5	Helianthemum nummularium	Helianthemum nummularium	atcctgttttacgagcaattcaaaaag gaatcaaaaatag	40	228034	3
6	Plantago	Plantago	atcctgtcttctcaaaataaaggttcag aaagcgagaaaaagg	43	3942	1
7	Esula	Esula	atcctattttccgaaaacaaaaaaagtt ccataaagacagaaaaaaag	49	124651	2

8	-	Nannochloropsis granulata	ctcatgaaagtg	12	240	4
9	-	Isoetes	attccgttgtttcatttcatgaacgaacgg	30	7092	1

Table 3-S3. Plant DNA sequences analyzed in the main text. MOTU names are the sequence names used in the analyses of the main text.

	Scientific name (alpine data base)	Rank (alpine database)	Scientific name (global database)	Rank (global database)	Resolved MOTU name	sequence	Sequence length
1	Asteraceae	family	-	no rank	Asteraceae MOTU 1	atcacgttttccgaaaacaaacaaagattcagaaagcgaaaatcaaaaag	50
2	Asteraceae	family	-	no rank	Asteraceae MOTU 2	atcacgttttccgaaaacaaaggttcagaaagcgaaaatcaaaaag	46
3	-	no rank	Scorzoneroides	genus	Leontodon	atcacgttttccgaaaacacacaaaggttcagaaagcgaaaataaaaaag	50
4	Asteraceae	family	-	no rank	Asteraceae MOTU 3	atcacgttttccgaaaacaaacaaaggttcagaaagcgaaaataaaaaag	50
5	Asteroideae	subfamily	-	no rank	Asteroideae MOTU 1	atcacgttttccgaaaacaaacaaaggttcagaaagcgaaaaaaaag	47
6	Asteroideae	subfamily	-	no rank	Asteroideae MOTU 2	atcacgttttccgaaaacaaacaaaggttcagaaagcgaaaagaaaaaaa	51

7	Alchemilla	genus	Alchemilla	genus	Alchemilla MOTU 1	atcccgttttatgaaaacaaaggtttcagaaagcgagaataaaataagg	48
8	Colurieae	tribe	-	no rank	Colurieae	atcccgttttatgaaaagaaacaagggtttcagaacgcgagaataaataaag	52
9	Alchemilla	genus	Alchemilla	genus	Alchemilla MOTU 2	atcccttatttatgaaaacaaaggtttcagaaagcgagaataaataaagg	49
10	Potentilla	genus	Potentilla	genus	Potentilla	atcccgttttatgaaaacaaacaagggtttcataaagcgagaataaataaag	52
11	Poeae	tribe	-	no rank	Poeae MOTU 1	atccgtgttttgagaaaacaaaggggttctcgaatcgaactataatacaaaggaaaag	58
12	Poeae	tribe	-	no rank	Poeae MOTU 2	atccgtgttttgagaaaacaaaggggttctcgaactagaatacaaaggaaaag	53
13	Cupressaceae	family	-	no rank	Cupressaceae	atccgatttctagagacaatagttttcctttccgagaacgg	40
14	Poeae	tribe	-	no rank	Poeae MOTU 3	atccgtgttttgagaaaacaaggaggttctcgaactagaatacaaaggaaaag	53
15	Myosotis	genus	Myosotis	genus	Myosotis	atccggttttccgaaaacaaaagttgaaaaagaaaaaaggaag	46

						gag	
16	Campanulaceae	family	-	no rank	Campanulaceae	atccggttttctgacaataacaaaaggttcagaaagcgaaaatcaaaaag	50
17	Boraginaceae	family	-	no rank	Boraginaceae	atccggttttccgaaaacaaaagttgaaaaagaaaaag	38
18	Poeae	tribe	-	no rank	Poeae MOTU 4	atccgtgttttgagaaggggttctcgaactagaatacaaaggaaaag	48
19	Viola tricolor	species	Viola	genus	Viola tricolor	atccggtttttttttaataaaaaaagtttatatagacagaataaaaaag	48
20	Ranunculus	genus	Ranunculus	genus	Ranunculus MOTU 1	atcctgctttcagaaagcaaaaaaagg	27
21	Ranunculus glacialis	species	Ranunculus glacialis	species	Ranunculus glacialis	atcctgctttcagaaaacaaaaaggggagttcagaaagcaaaaatagg	48
22	Veronica	genus	Veronica	subgenus	Veronica	atcctgtcttctaaaaacaaaggttcagaaagtgaaaaaaaaagg	45
23	Bartsia alpina	species	Bartsia alpina	species	Bartsia alpina	atcctgttttcttaaaacaaagattcagaaaacgaaaaataaag	44
24	-	no rank	Saxifraga kotschyi	species	Saxifraga MOTU 1	atcctgttttaagaaagcaaaacacaagggttcagaaaaagcgaataaaaaaaaaag	57

25	Solanaceae	family	-	no rank	cf. Solanaceae	atcctgttttccgaaaacaaacaaaggttcagaaaaaaag	40
26	Apiaceae	family	-	no rank	Apiaceae	atcctattttccaaaaacaaacaaaggcccagaaggtgaaaaaag	45
27	Plantago	genus	Plantago	genus	Plantago	atcctgtcttctccaaataaaggttcagaaagcgaaaaggg	41
28	Trifolium	genus	Trifolium	genus	Trifolium MOTU 1	atccttctttccgaaaacaaaataaaagttcagaaagttaaaataaaaaagg	52
29	Betulaceae	family	-	no rank	Betulaceae	atcctgttttccgaaaacaaataaaacaaatttaagggttcataaagtgagaataaaaaag	61
30	Saxifraga	genus	Saxifraga	genus	Saxifraga MOTU 2	atcctgttttaagaaagcaaaacacaagggttcagaaaaagcgaataaaaaaaaaaag	58
31	-	no rank	Lotus	genus	Lotus	atcctgctttacgaaaacaagggaaagttcagttaagaaagcgacgagaaaaatg	55
32	Saxifraga oppositifolia	species	Saxifraga	genus	Saxifraga oppositifolia	atcctgttttaagaaagcaaaacacaagggttcagaaaaagcgataaaaaaaag	55
33	Primulaceae	family	-	no rank	Primulaceae	atcctcttttagaaaacaaaaatgaaaggaaaataaaaagggg	44

34	Vaccinium uliginosum	species	Vaccinium	genus	Vaccinium uliginosum	atccttctttttggcaaacaaataaagattccgaaagaaataagg	45
35	-	no rank	Ranunculus	genus	Ranunculus MOTU 2	atcctgctttcagaaaacaaaaaggggggttcagaaagaaaaaaagg	48
36	Soldanella	genus	Soldanella	genus	Soldanella	atcctctttttagaaaacaaagtttaagtaaactaaaaag	41
37	Ranunculaceae	family	-	no rank	Ranunculaceae	atcctgttttcagaaaacaaaaagggttcagaaagcaagaatcaaaaag	49
38	-	no rank	Melica	genus	Melica	atccttgttttgagaaaacaagtggttctcgaactagaatccaaaggaaaag	52
39	Trifolium	genus	Trifolium	genus	Trifolium MOTU 2	atccttctttctgaaaacaaataaaagttcagaaagttaaaatcaaaaagg	51
40	Euphrasia	genus	Euphrasia	genus	Euphrasia	atcctgtttttttccaaacaaagattcagaaaacgaaaaaaag	42
41	Mentheae	tribe	-	no rank	Mentheae	atcctgttttctcaaaacaaag	22
42	Onagraceae	family	-	no rank	Onagraceae	atcctattttacgaaaaccagcaccgggggtttagaaagcgatcaaaaaaaag	53

43	Trifolium pallescens	species	Trifolium pallescens	species	Trifolium pallescens	atccttctttccgaaaacaaaataaaagttcagaaagttaaaataagg	48
44	Brassicaceae	family	-	no rank	Brassicaceae	atcctggtttacgcgcacaaaccggagtttaaaaagcgagaaaaagg	47
45	Sempervivum tectorum	species	Sempervivum tectorum	species	Sempervivum tectorum	atcttcttttccgaaaccaaacccaactacgggggttcaaaaagcgaaaataaaaaag	58
46	-	no rank	Crassula	genus	Crassula	atcttcttttccaaaaccaaacccactcaaggggttaacaaagcgaaaaaggg	54
47	-	no rank	Luzula	genus	Luzula	atcttaatctggagaaaatttgtttctctataaaaactagaatcaaaaag	50
48	-	no rank	Pedicularis attollens	species	Pedicularis	atcttctttttttttcaaaacaaaggtttagaaaacgaaaaaaaag	46
49	Carex	genus	Carex	genus	Carex MOTU 1	atcttctttttgagaaaaagaaatatataaaatatttcttatttcagataagaaataatatttttttcttatctaatattaaa	82
50	Carex	genus	Carex	genus	Carex MOTU 2	atcttctttttgagaaaaagaaatatataaaatatttcttatttcagataagaaagaataatattttcttatctaatattaaa	83

51	Oxyria digyna	species	Oxyria digyna	species	Oxyria digyna	ctccttcgtccaaaaggaagaataaaaaag	30
52	Rumex	genus	Rumex	genus	Rumex	ctcctcctttccaaaaggaagaataaaaaag	31
53	Caryophyllaceae	family	-	no rank	Caryophyllaceae	ctccttaattgtctcctaaaaaaaacggattcaaaaagaaagaataaaaaag	52
54	Cerastium cerastoides	species	Cerastium cerastoides	species	Cerastium cerastoides	ctccttaatagtctcctaaaaaaaacggattcaaaaagaataaaaaag	48
55	Silene acaulis	species	Silene acaulis	species	Silene acaulis	ctccttttatttttcctttttttttcaaaagcaaagatgaagaaagcaagaaaaaaaagg	57
56	Cardamineae	tribe	-	no rank	Cardamineae	atccttgtttacgcgaacaaaccggagtttagaaagcgagaaaaaagg	48

Table 3-S4. Parameter estimates of the site occupancy model on Domestic animals. HDPI: High Density Probability Interval.

	p_{10}	p_{10} HPDI	p_{11}	p_{11} HPDI	Ψ	Ψ HPDI
Ovis	0.010	0.002-0.021	0.168	0.062-0.358	0.114	0.021-0.292
Bos	0.007	0.002-0.015	0.201	0.075-0.372	0.083	0.021-0.202

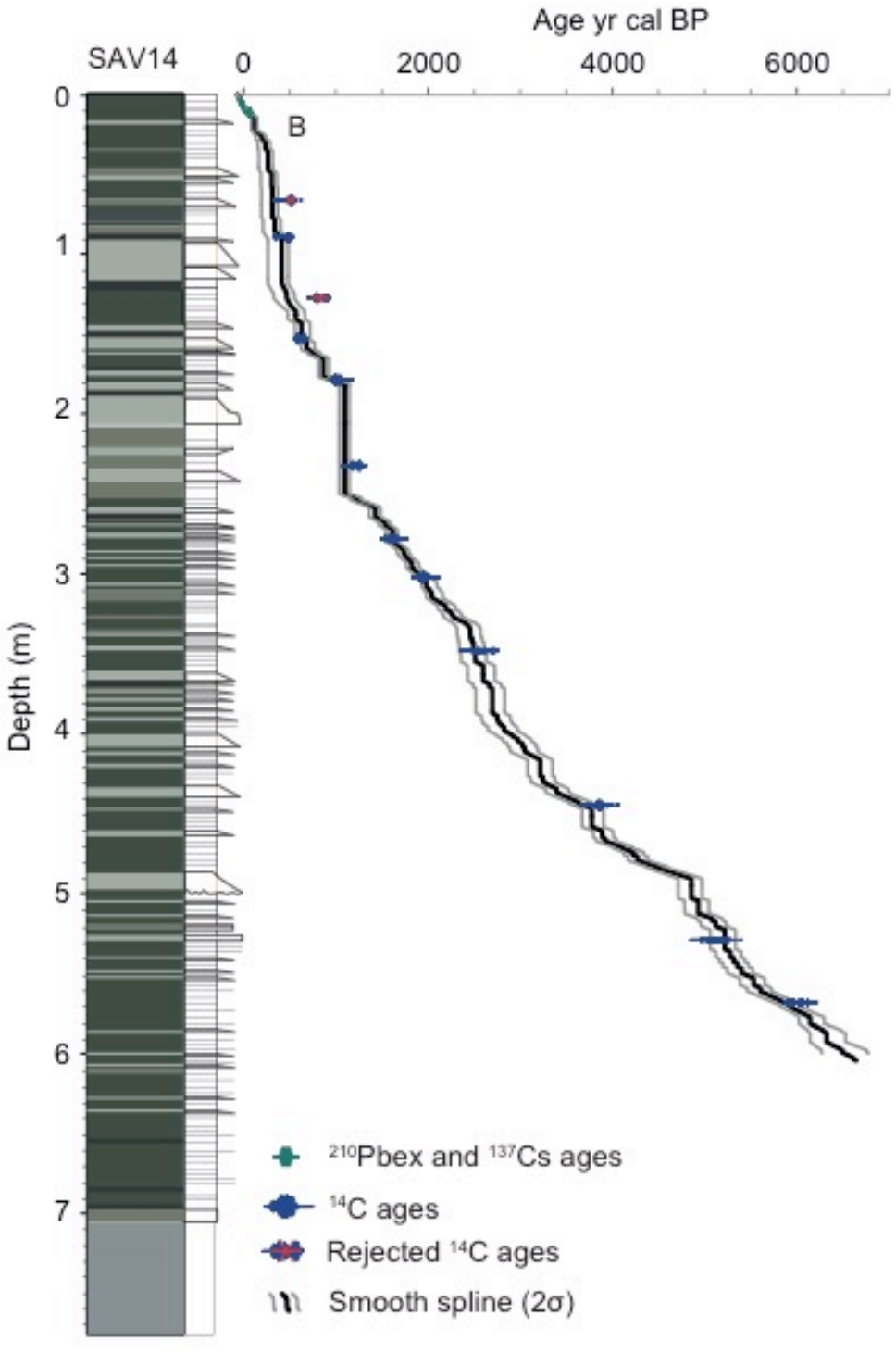
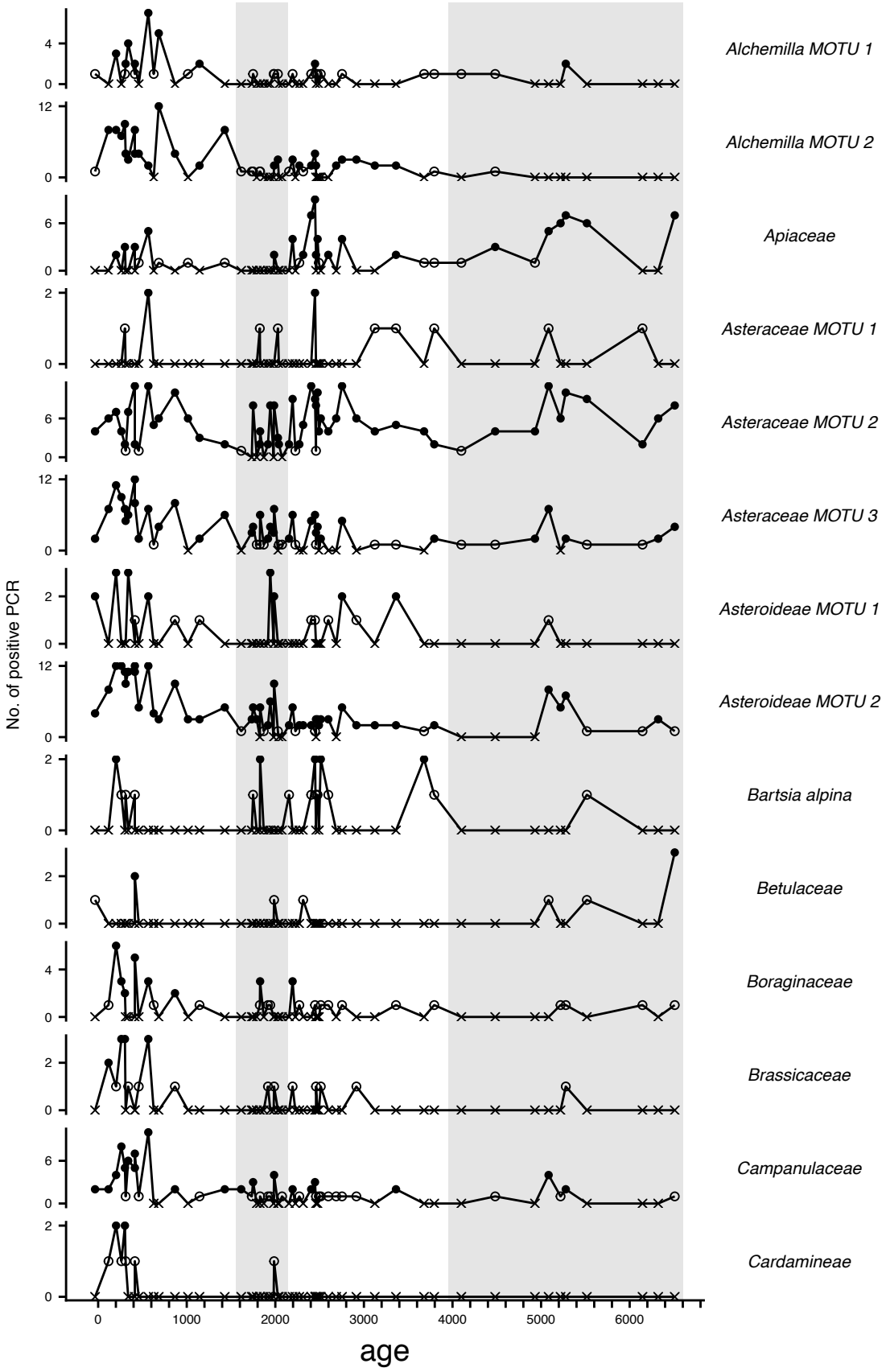
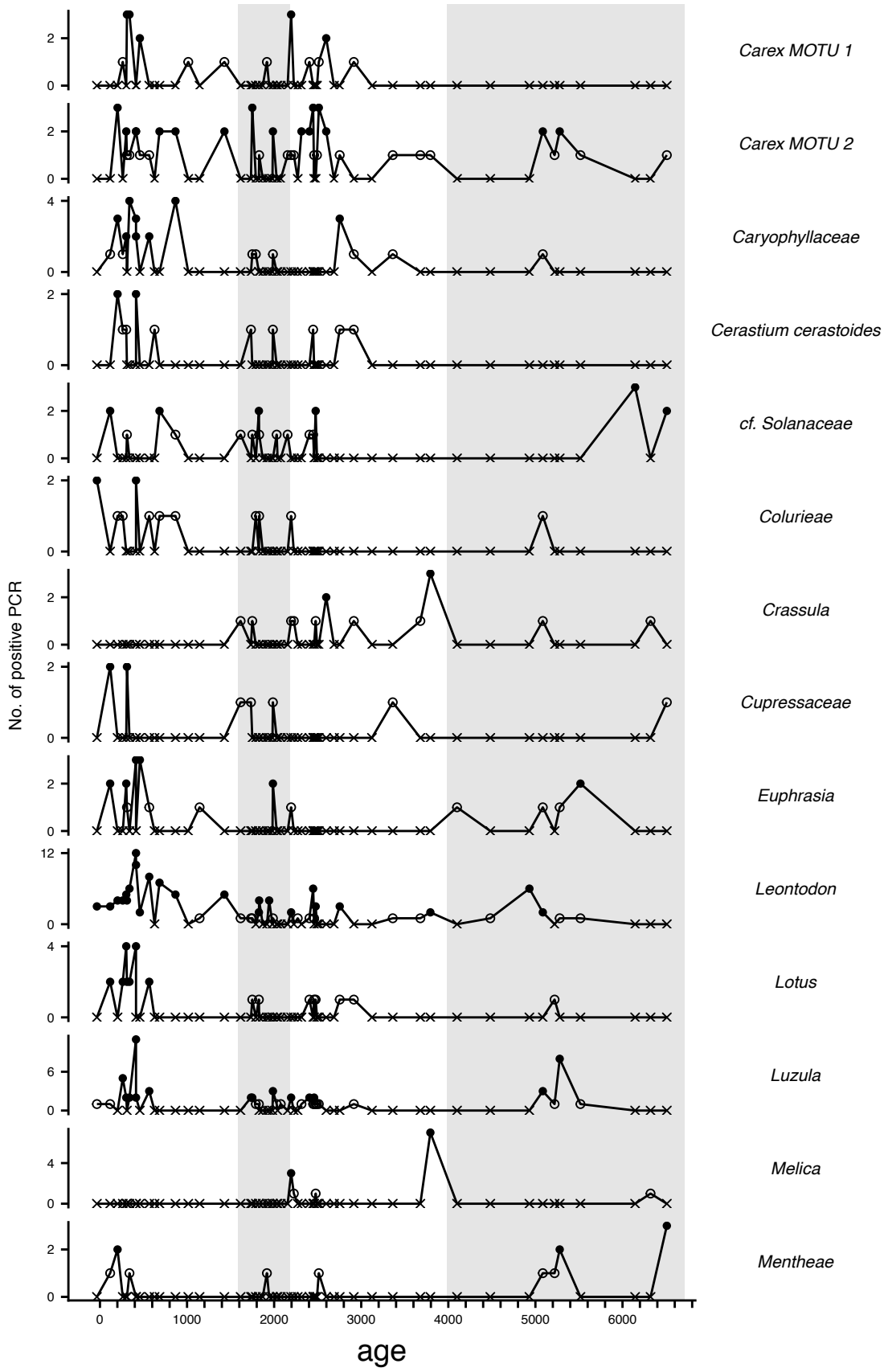
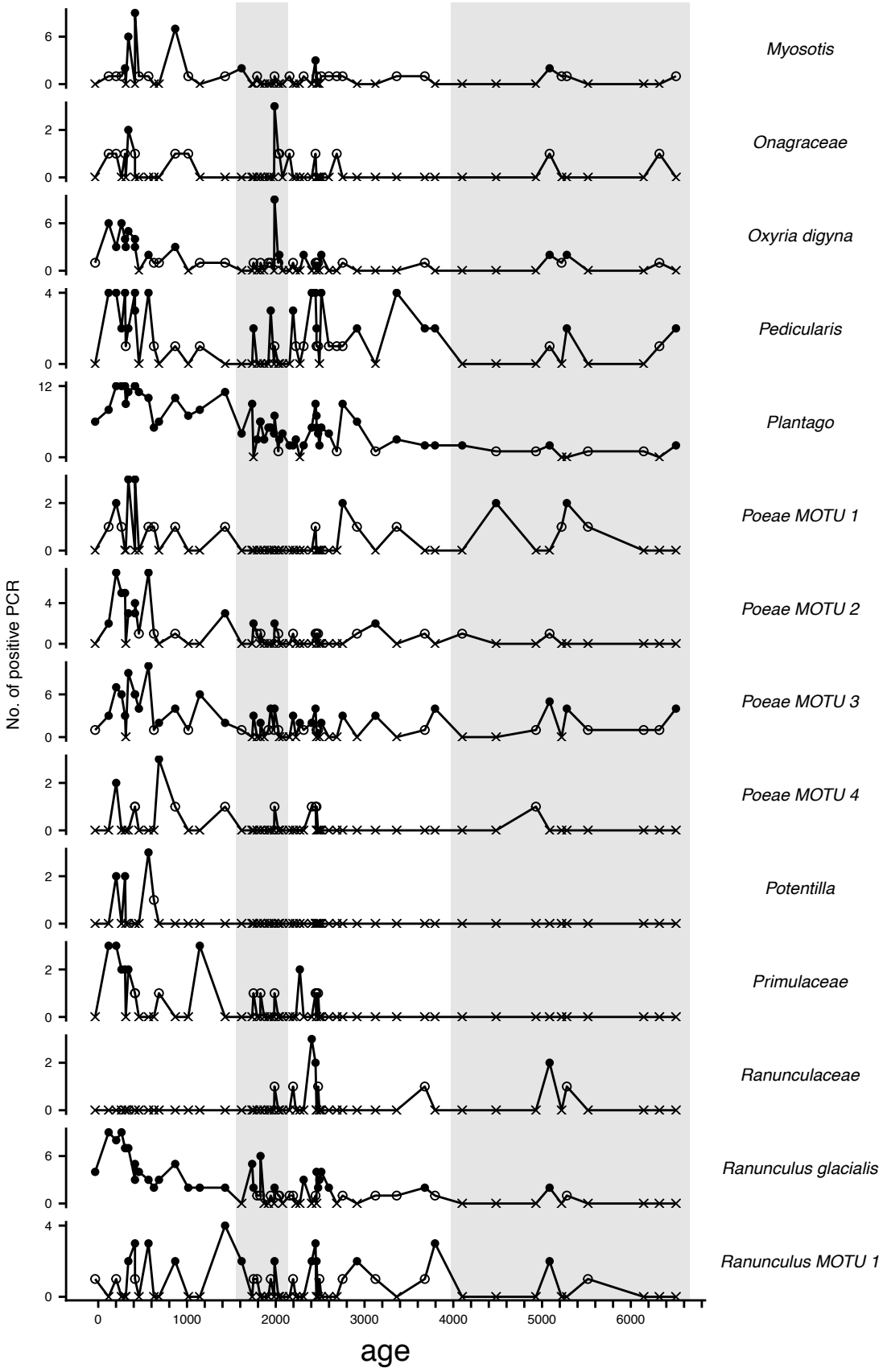


Figure 3-S1 Age model of the sediment core.

Figure 3-S2 Chronosequences of number of positive PCR replicates of all plant taxa analyzed. Time intervals of Periods B and D are shaded.


No. of positive PCR
Alchemilla MOTU 1
Alchemilla MOTU 2
Apiaceae
Asteraceae MOTU 1
Asteraceae MOTU 2
Asteraceae MOTU 3
Asteroideae MOTU 1
Asteroideae MOTU 2
Bartsia alpina
Betulaceae
Boraginaceae
Brassicaceae
Campanulaceae
Cardamineae
0
1000
2000
3000
4000
5000
6000
age


Carex MOTU 1
Carex MOTU 2
Caryophyllaceae
Cerastium cerastoides
cf. Solanaceae
Colurieae
Crassula
Cupressaceae
Euphrasia
Leontodon
Lotus
Luzula
Melica
Mentheae
No. of positive PCR
age
0
1000
2000
3000
4000
5000
6000


No. of positive PCR
Myosotis
Onagraceae
Oxyria digyna
Pedicularis
Plantago
Poeae MOTU 1
Poeae MOTU 2
Poeae MOTU 3
Poeae MOTU 4
Potentilla
Primulaceae
Ranunculaceae
Ranunculus glacialis
Ranunculus MOTU 1
0
1000
2000
3000
4000
5000
6000
age

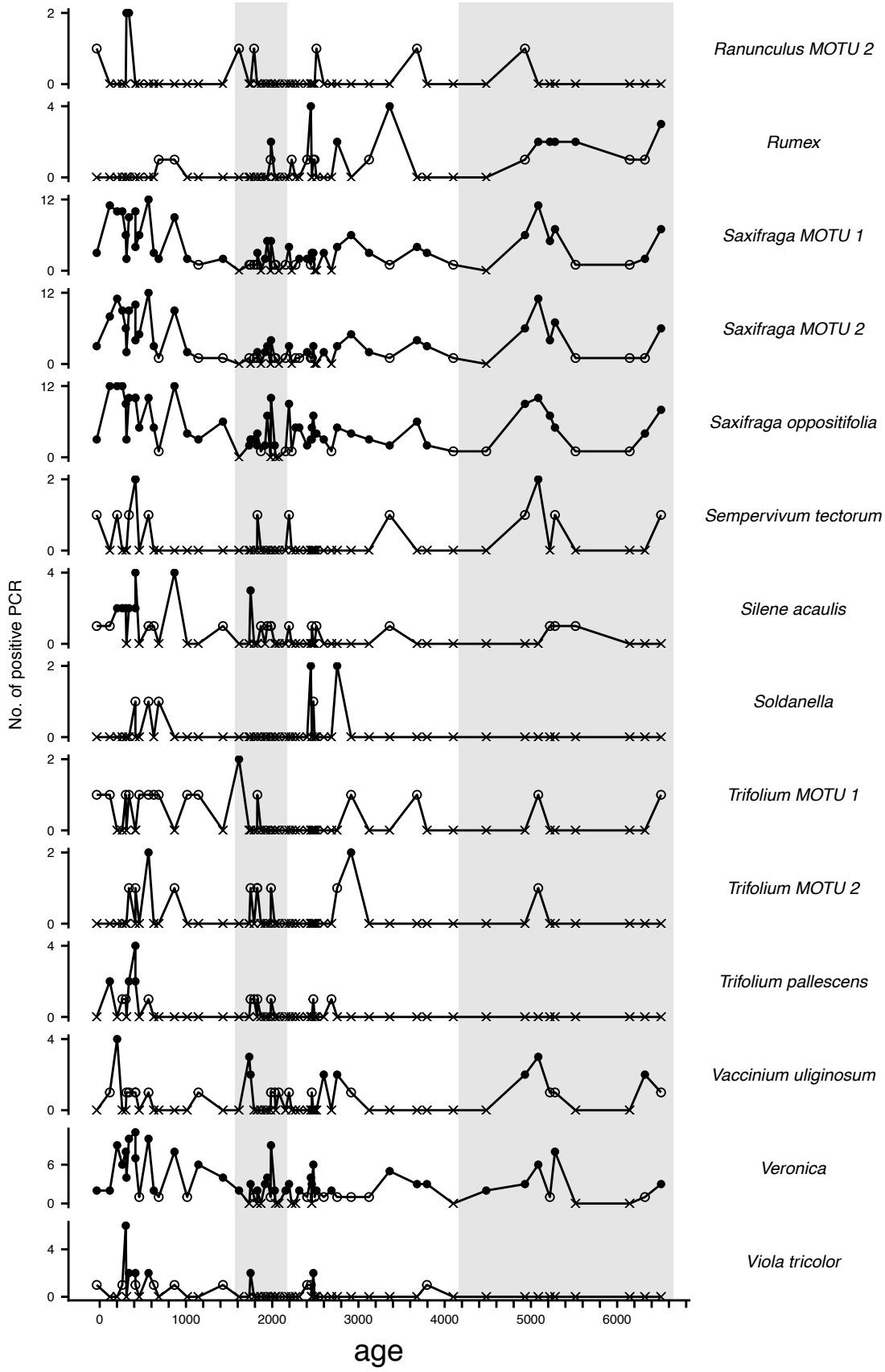

No. of positive PCR
Ranunculus MOTU 2
Rumex
Saxifraga MOTU 1
Saxifraga MOTU 2
Saxifraga oppositifolia
Sempervivum tectorum
Silene acaulis
Soldanella
Trifolium MOTU 1
Trifolium MOTU 2
Trifolium pallescens
Vaccinium uliginosum
Veronica
Viola tricolor
0
1000
2000
3000
4000
5000
6000
age

Figure 3-S3 PCA plots of the DNA samples of Period B. based on the presence/absence matrix of species-site. Numbers in red represent sample number. The first two principal components (PC1 and PC2) bear 29.9% and 12.8% of the total variance, respectively.

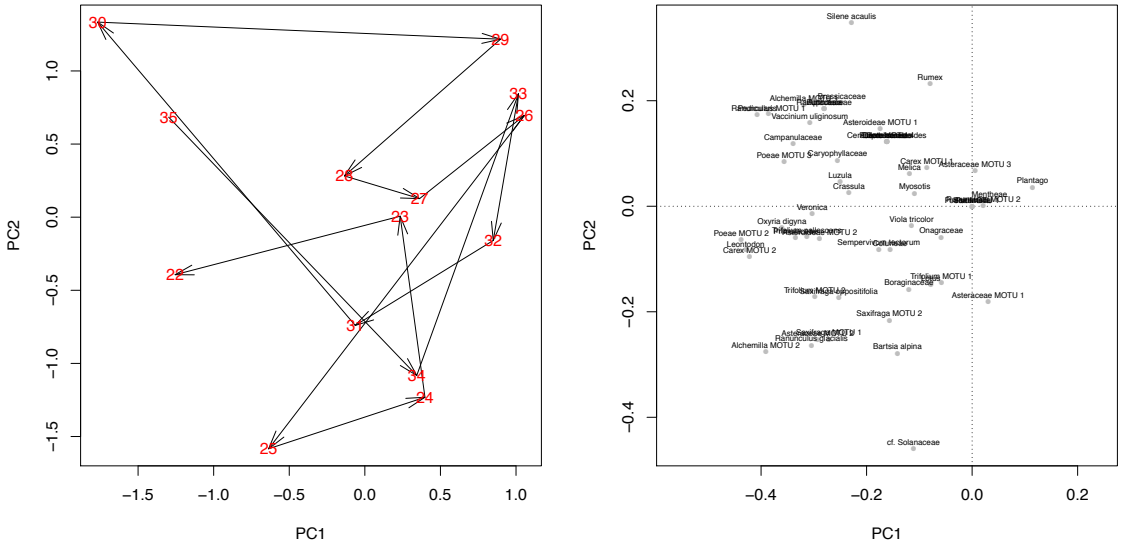


Figure 3-S4 Granulometric of sediment samples. Time intervals of Periods B and D are shaded.

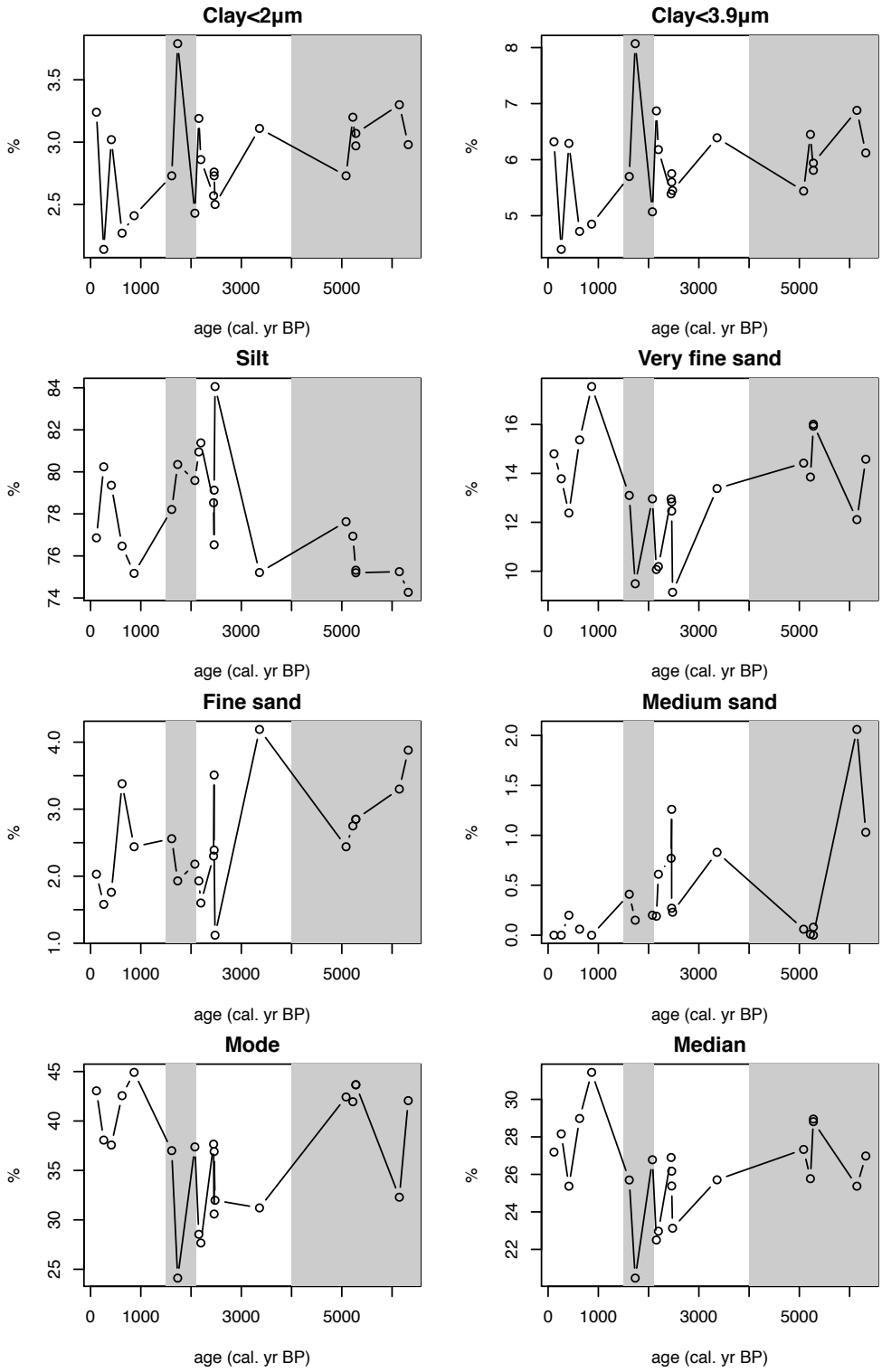
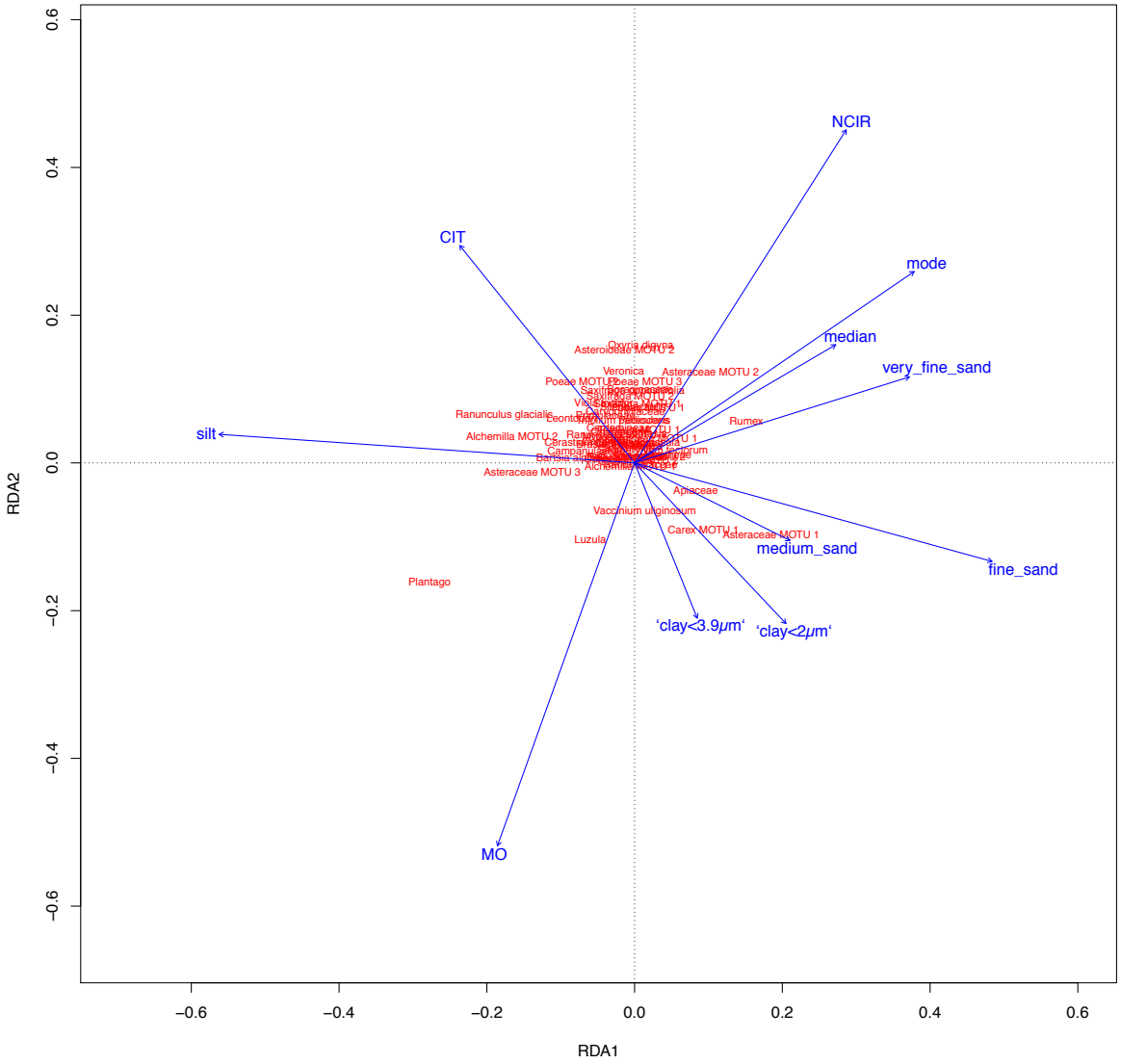


Figure 3-S5 Species plot of the first two axes of RDA with arrows representing scores of sample granulometric parameters. Organic matter (OM). carbonate ignous residue (CIT) and non-carbonate igneous residue (NCIT). The RDA is based on Hellinger-transformed table of number of positive PCR replicates. Only OM is significant (p=0.024. 999 permutations).



3.10.2 Testing the impact of flood deposit content on DNA detection

The sediment samples from which the DNA analyzed in the main text was extracted contained different proportions of flood deposits (see Table S1). We performed several tests to evaluate the impact of the variability of flood deposit content on plant DNA detection and, eventually, on the ecological analyses presented in the main text.

1) We checked the influence of flood deposit percentages in sediment samples on the total DNA detection. We used linear regression to assess the relationships between DNA detection (total number of positive PCR replicates and total number of presence, i.e. total number of taxa with positive PCR replicates) and the percentages of flood deposit, (identified by eye, see Materials and Methods in main text) while taking into account the sampling period (Table S5). Both total numbers of positive replicates and of presence were significantly different among the four periods. The numbers of positive PCR replicates showed a significant but weak relationship with flood deposit percentage ($p=0.024$), while the relationship between flood deposits and total number of presence was not significant ($p=0.084$).

2) We checked the influence of flood deposit percentages on the RDA performed in the main text. To this end, we re-run the analysis by conditioning the percentage of flood deposit. The results were nearly identical, with the proportion of variance explained by temperature being slightly smaller in the conditioned RDA than in the unconditioned one (Table S6).

3) Samples with more than 70% flood deposit tended to have more plant DNA detected (Figure S6), making them potential sources of compositional variability. To test if flood deposit content can bias multivariate dispersion tests, we re-run the tests on a dataset without samples containing more than 70% flood deposit. The test results on the reduced dataset were very close to the ones on the full dataset. Particularly, tests on both datasets showed that the dispersion level of Period B was significantly higher than those of Periods A and C, and the respectively p-values were close. On the other hand, the levels of dispersion in Periods A and C were significantly different in the test on the full dataset ($p=0.0073$), but not so in the test on the reduced dataset (Table S7).

In summary, we found that the proportion of flood deposit in sediment samples had mild effects on the detection rates of plant DNA, and that our tests on plant community composition were robust under these effects. The mechanisms of these effects are unknown and deserve further investigations, but they are out of the scope of the present study.

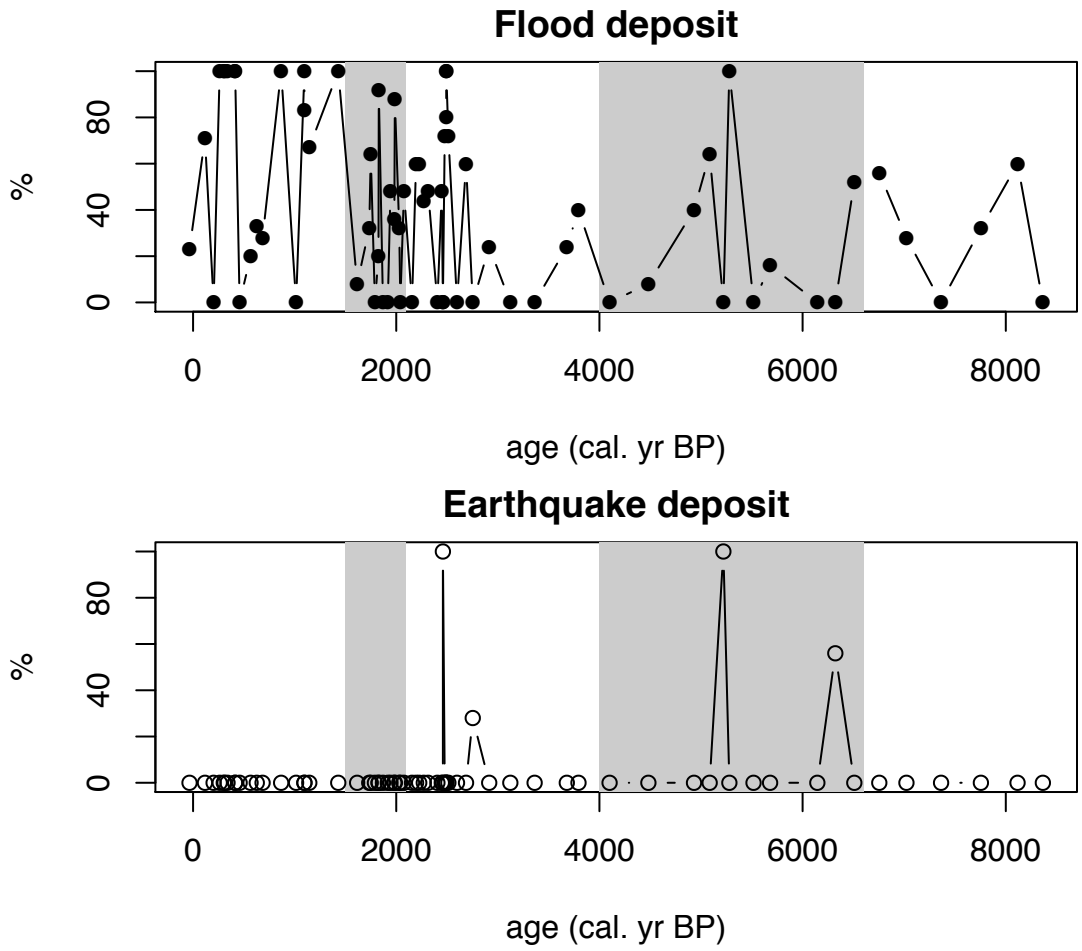


Figure 3-S6 Percentages of deposits of flood and earthquake events in sediment samples. Time intervals of Periods B and D are shaded.

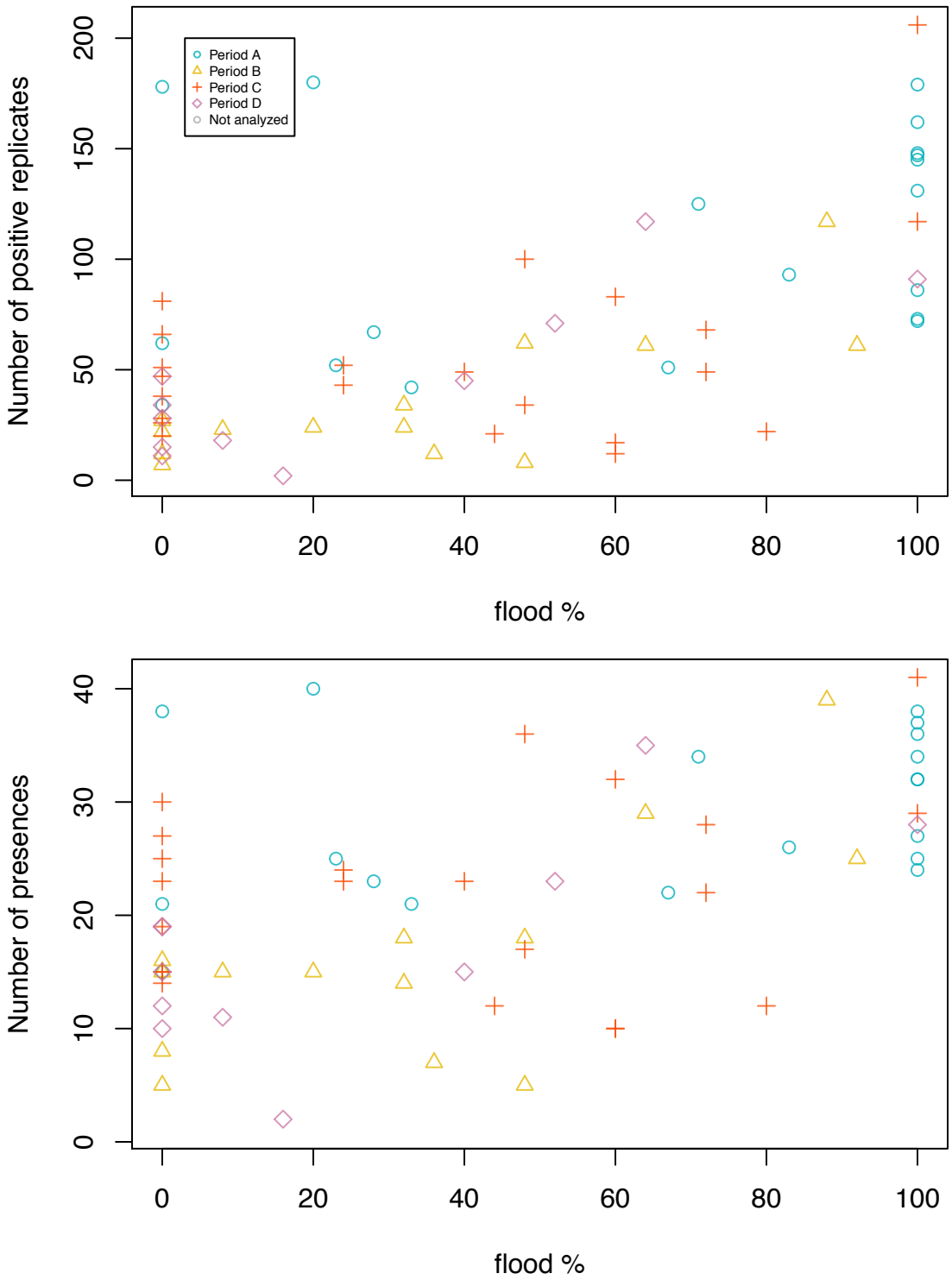


Figure 3-S7 Numbers of total positive replicates and of total presences with sample flood deposit content (in percentage).

	Total number of positive replicates			Total number of taxa presence		
	Estimate	*CI*	*p*	*Estimate*	*CI*	*p*
(Intercept)	90.65	60.08 – 121.21	<.001***	27.09	20.21 – 33.97	<.001***
% flood	0.48	0.07 – 0.90	.024*	0.08	-0.01 – 0.18	.084
Period B	-70.86	-112.51 – -29.22	.001***	-16.30	-25.67 – -6.93	.001***
Period C	-38.65	-77.81 – 0.51	.053	-4.11	-12.93 – 4.70	.353
Period D	-64.45	-105.93 – -22.97	.003**	-12.12	-21.46 – -2.79	.012**
% flood:period B	0.14	-0.01-0.29	.0.06	0.10	-0.09 – 0.28	.298
% flood:period C	-0.46	-1.15 – 0.23	.189	-0.10	-0.25 – 0.06	.217
% flood:period D	0.37	-0.41 – 1.14	.344	0.09	-0.08 – 0.27	.286
Observations	59			59		
R^2 / adj. R^2	.581 / .523			.491 / .421		

Table 3-S5. Linear regressions Sample total number of positive replicates and Sample total number of taxa presence of with flood deposit proportion (% flood). ***: <=0.001, **: <0.01, * <0.05.

Test	p-value (Temperature)	% variance explained (Temperature)	p-value (Pastoralism)	% variance explained p-value (Temperature)
Conditioned	0.0083	3.21	0.7387	1.38
Unconditioned	0.0053	3.73	0.7814	1.27

Table 3-S6. Comparison between RDAs conditioned and unconditioned with flood deposit percentage. P-values were obtained by 9999 permutations.

	Reduced		Full	
Period pair	$F_{\text{permutation}}$	$P_{permutation}$	$F_{\text{permutation}}$	$P_{permutation}$
A-B	-3.66	0.0018	-4.80	0.0002
A-C	-1.57	0.1317	-2.97	0.0047
A-D	-2.72	0.0166	-3.38	0.0010
B-C	2.77	0.0108	2.57	0.0165
B-D	1.72	0.1034	1.65	0.1151
C-D	-0.83	0.4224	-0.54	0.5999

Table 3-S7. Comparison between the results of multivariate dispersion tests on the full plant DNA dataset (Full) and on the dataset containing only the samples with less than 70% flood deposit (Reduced).

Chapter 4. Lake sediment DNA reveals complex effects of climate, land use and vegetation on soil erosion in a high mountain environment (Article C)

Wentao Chen[1], Manon Bajard[2,3], Pierre Sabatier[2], Jérôme Poulenard[2], Fabien Arnaud[2], David Etienne[4], Delphine Rioux[1], Pierre Taberlet[1], Jacque-Louis de Beaulieu[5], Gentile Francesco Ficetola[1,5]

[1] Univesité Grenoble Alpes, CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

[2] Environment Dynamics and Territories of Mountains (EDYTEM), Université Savoie Mont Blanc, CNRS, 73000 Chambéry, France

[3]Centre for Earth Evolution and Dynamics (CEED), University of Oslo, PO Box 1047, Blindern, NO-0316, Oslo, Norway

[4] UMR INRA 42 CARRTEL, Université Savoie Mont Blanc, 73376 Le Bourget-du-Lac, France

[5] Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (UMR 7263 IMBE), Aix-Marseille Université, 13100, Aix-en-Provence, France

[6] Department of Environmental Science and Policy, Università degli Studi di Milano. Via Celoria 26, 20133 Milano, Italy

Keywords:

Abstract

Soil erosion is a crucial threat to ecosystem functioning. Livestock farming and climate change are two major drivers of soil loss in high-altitude mountain areas. Previous studies on this topic usually considered these drivers separately, and rarely looked into the complex interactions among human activities, climate, vegetation, and soil erosion. Here we present a multiproxy reconstruction of the long-term environmental history in an alpine area through the last six thousand years, by numerically analyzing environmental DNA and coprophilous fungus spores extracted from a lake sediment core, combined with long-term total solar irradiance records as climate indicator. We found a major change of vegetation from pine forest to tall-forb grassland and meadows, accompanied by an increasing soil erosion rate, after the introduction of livestock into the area in ca. 2700 cal. yr BP. This transition resulted from livestock farming was also characterized by a significant increase in the sensitivity of soil erosion rate to climate variability. Sedimentary DNA combined with traditional environmental proxies allow to unravel complex interactions of human activities, climate, vegetation and soil through an integrated temporal analysis, providing excellent opportunities to better understand the mechanisms that drive the change of ecosystem functioning.

4.1 Introduction

Human activities have significantly altered the environment for millennia, with increasing impacts on many ecosystem functions, ranging from biogeochemical cycles to productivity (Vitousek et al., 1997). Such impacts are often mediated by changes in land-use (Metzger, Rounsevell, Acosta-Michlik, Leemans, & Schröter, 2006) and in biodiversity (Cardinale et al., 2012). Both land-use and biodiversity changes can affect the resilience of ecosystem functioning faced to anthropogenic or climatic disturbances (Dearing, 2008; Folke, 2006; Loreau & de Mazancourt, 2013; Peterson, Allen, & Holling, 1998). Soil is a major provider of essential ecosystem functions and services (Barrios, 2007; Fitter et al., 2005). How human activities impact soil resistance to erosion is a crucial issue to the resilience of ecosystem functioning. Vegetation can reduce erosion by stabilizing and protecting soil from wind and water runoff (Zuazo & Pleguezuelo, 2008). Experiments and models have shown that the ability of plants to prevent soil erosion depends on both vegetation type and plant species diversity (Berendse, van Ruijven, Jongejans, & Keesstra, 2015; Zuazo & Pleguezuelo, 2008). Human activities can therefore trigger soil erosion directly by overgrazing of livestock (Trimble & Mendel, 1995), and indirectly by modifying plant communities through land-use change (Bajard et al., 2015; Bajard, Poulenard, Sabatier, Etienne, et al., 2017; Giguet-Covex et al., 2011, 2014). On the other hand, humans can directly alter soil susceptibility to erosion via changes in land management (Sabatier et al., 2014; Vanwalleghem et al., 2017).

Observational and experimental studies have provided insights on the linkages between land-use change, vegetation and soil erosion (e.g. Ludwig, Wilcox, Breshears, Tongway, & Imeson, 2005; Martin, Pohl, Alewell, Körner, & Rixen, 2010; Mohammad & Adam, 2010), and between on the impacts of climate change (Arnaud et al., 2016; Herman et al., 2013). However these studies generally had limitations such as a limited temporal extent, and the lack of integration of multiple drivers of erosion, including climate, vegetation change and

other human stressors, thus have a limited power to detect long-term causal links. A better understanding on the drivers of soil degradation and their consequences for ecosystem functioning requires considering soil erosion within a historical context, and assessing the potential impact of multiple, interacting drivers (Bajard, Poulenard, Sabatier, Etienne, et al., 2017; Dotterweich, 2013; Vanwalleghem et al., 2017).

Retro-observational studies based on lake sediment records provide crucial knowledge about the processes through which past human activities affect vegetation and soil, and on the complexity of drivers. For instance, in high mountain environments, human-induced deforestation and livestock grazing are major causes of vegetation changes (Pansu, Giguet-Covex, et al., 2015; Pini et al., 2017). These changes are often accompanied by increased erosion rates, usually inferred from increased flood deposit magnitude and frequency (Brisset et al., 2017; Giguet-Covex et al., 2011). Moreover, overgrazing by livestock can also directly contribute soil erosion by compacting soils, reducing water infiltration and detaching soil particles (Trimble & Mendel, 1995). Soil erosion is a direct result of wind or water flows, thus climate variabilities that control precipitation and storms must play an important role in the erosion process (Arnaud et al., 2016; Bärring, Jönsson, Mattsson, & Åhman, 2003; Nearing, Pruski, & O'Neal, 2004; Sabatier et al., 2012). However, to date, only a few studies have considered the complex links between land use change and soil susceptibility to climatic or anthropogenic disturbances (e.g. Bajard, Poulenard, Sabatier, Develle, et al., 2017; Dearing, 2008), despite appropriate analytical approaches are now able to disentangle such causal links (Shipley, 2000b).

In the Northern Alps, early human-induced extensive deforestations and livestock grazing at high altitudes were recorded around 3000-3600 yr BP, corresponding to the Middle Bronze age [e.g. (Giguet-Covex et al., 2011) for the Northern French Alps, (Röpke, Stobbe, Oeggl, Kalis, & Tinner, 2011) for the Swiss Alps and (Dietre et al., 2014) for the Austrian Alps], with subsequent intensifications during the Roman period (around 2000-1600 yr BP) and the Late Medieval – Modern periods (Bajard, Poulenard, Sabatier, Develle, et al., 2017; Pansu, Giguet-Covex, et al., 2015). These major periods of strong human impacts were

often accompanied by a rise in soil erosion at local scales (Arnaud et al., 2016), but climatic conditions showed strong variability across them (Büntgen et al., 2011; Giguet-Covex et al., 2012), thus making it challenging the identification of drivers of erosion. Such complex history provides excellent opportunities to test the relative influences of climate variability, human activities and vegetation on soil erosion, and eventually to disentangle the complex interactions among them.

Sedimentary DNA (SedDNA) is an emerging tool to reconstruct long-term environment trajectories, providing exceptional information of the evolution of past communities (Alsos, Sjögren, et al., 2016b; Ficetola, Poulenard, et al., 2018; Giguet-Covex et al., 2014; Pansu, De Danieli, et al., 2015). In contrast to pollen analysis, which provides regional assessments of vegetation, SedDNA provides a more localized view capturing vegetation development in small lake catchments (Edwards et al., 2018; Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, Pedersen, et al., 2017). SedDNA can be combined with other paleorecords such as pollen, non-pollen palynomorphs (NPPs), and also geochemical/sedimentological data, in order to obtain highly valuable information about paleoenvironments (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). Based on SedDNA, we reconstruct here the long-term links between human activities, vegetation change and erosion in a subalpine lake catchment in the Alps (Figure 4-1).

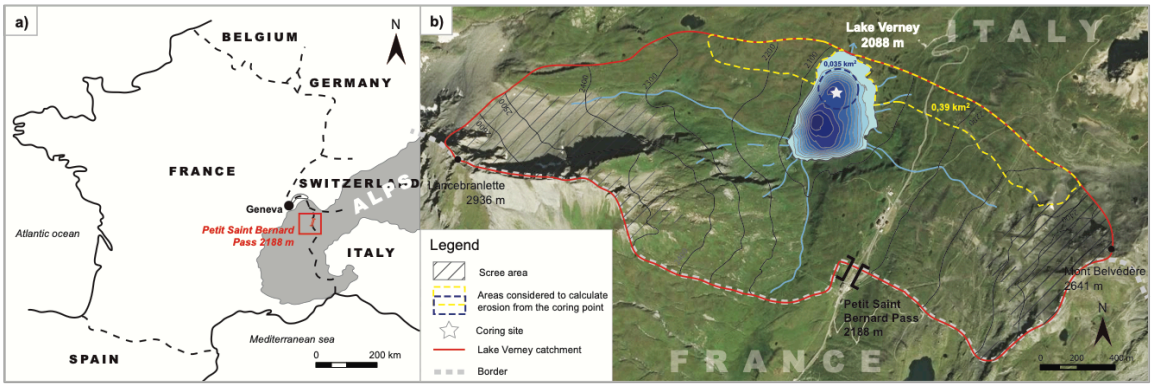


Figure 4-1: Location of Petit Saint Bernard Pass (a), pictures of the Lake Verney catchment (b), locations of sampling sites and a simplified geologic map covering the area of the pass (c). Abbreviations: LTH = La Thuile.

Here we provide an unprecedented analysis on how human activities, alteration of plant communities and climate change determined soil erosion through more than five millennia. To assess the direct and indirect effects of these processes on soil erosion, we combined information on soil degradation with the sedimentary DNA of animal and plant communities, spores of coprophilous fungi and paleoclimatic records in the literature. Our study provides valuable insights on the complex mechanisms of anthropogenic impacts over ecosystem functioning along mid- to late Holocene.

4.2 Results

4.2.1 Sedimentology and chronology

The sedimentology of Lake Verney sequence, including its description, and analyses of Loss On Ignition, grain-size and XRF geochemistry, was detailed in Bajard, Poulenard, Sabatier, Etienne, et al., (2017). The grain-size in constant through most of the 2.4 m of the sequence. Only the bottom of the sequence (Unit 4 in Figure 4-S1) displays a complex and different sedimentology. Changes in geochemistry throughout the sequence, as the decrease in carbonates contents, mainly reflects evolution of soils in the catchment and their acidification.

The age-depth model of Lake Verney sediment sequence was modified from Bajard, Poulenard, Sabatier, Etienne, et al., (2017) to include pollen-derived age. Two dates, 2100 yrs BP and 5050 yrs BP, were added, reflecting first known appearance of pollen of *Juglans* and *Castanea* found at 0.95 m, and *Picea* at 1.68 m, respectively, in this mountain area. Considering these new chronomarkers, the date at 1.05 m was discarded and the date at 1.21 m was re-integrated to the model (Figure 4-S1). The bottom of the sequence corresponding to units 3 and 4 displays an uncertain chronology and was therefore not considered in this study. The uppermost 1.7 m of sediment, corresponding to units 1 and 2, spans the last 6000 yr and the mean sedimentation rate is 0.4 $mm.yr^{-1}$.

4.2.2 DNA dataset

After filtering, the final dataset consisted of 54 sediment samples and contained two mammals taxa (*Bos* sp. and *Ovis* sp., i.e. cattle and sheep) and 81 plant molecular taxonomic units (MOTUs), of which 16 were assigned at the species level, 31 at the genus and 34 at the family/subfamily/tribe level. The DNA of domestic animals was mostly detected after 2500 cal. yr BP, with a particularly high frequency between 400 and 1000 cal. yr BP. Pines (*Pinus*) DNA was mostly detected before its nearly disappearance in ~2600 cal. yr BP, and partially recovered between ~1400 and 1000 cal. yr BP (Fig. S_STR).

4.2.3 Plant community variation

We applied Multivariate regression tree (MRT) analysis on the plant DNA dataset to test if the development of vegetation was structured in distinct phases. The MRT analysis identified a three-leaf tree, suggesting a three-phase grouping for the plant DNA dataset: Phase A with samples older than 2620 cal. yr BP (18 samples), Phase B for the period between 1030 and 2620 cal. yr BP (15 samples), and Phase C representing the period younger than 1030 cal. yr BP (21 samples; Figure 4-2 and Figure 4-S2). The analysis of similarity (ANOSIM) confirmed that the differences among such groups were significant (permutation tests: R^2=0.551, P=0.0001 between Phases A and B, and R^2=0.603, P=0.0001 between Phases B and C).

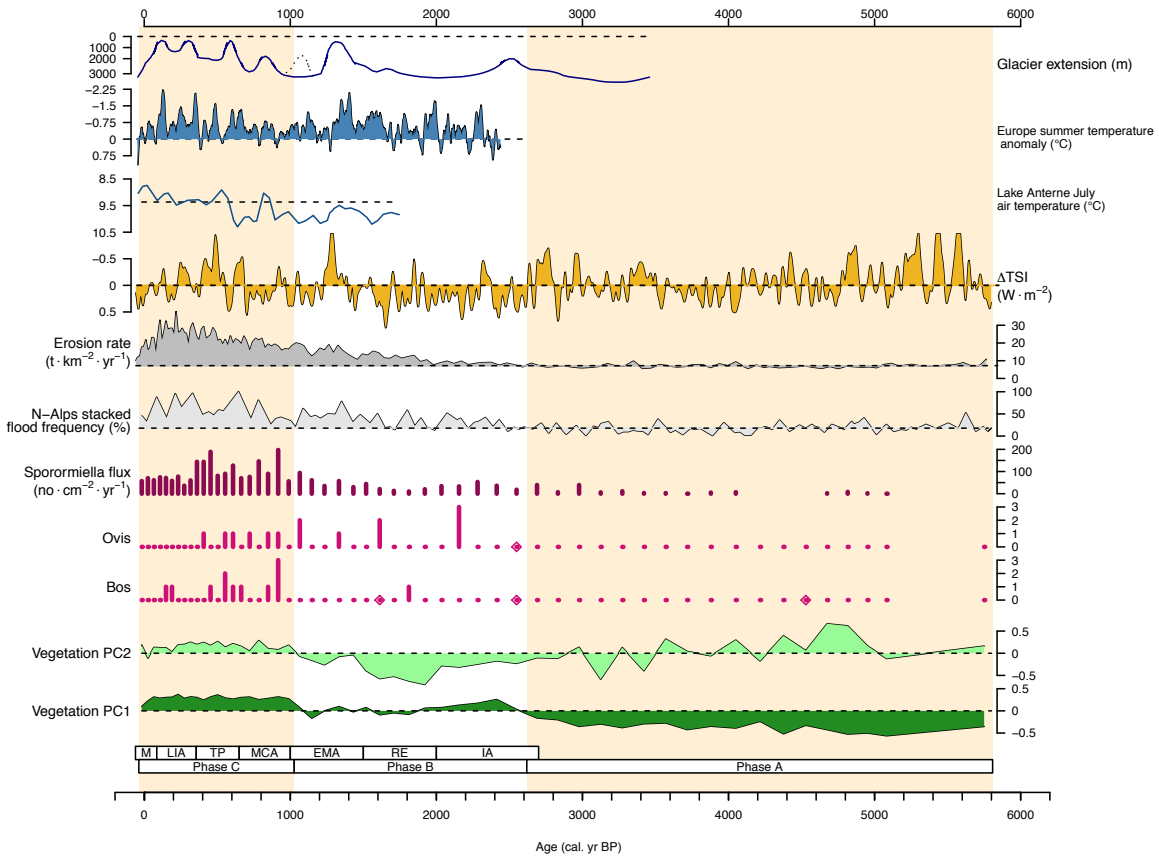


Figure 4-2: Times series of (from bottom to top) plant community change (the first two principle components of vegetation, PC1 and PC2), positive PCR replication number of cattle (Bos) and sheep (Ovis) DNA, coprophilous fungi spore accumulation rate, Northern Alps stacked flood frequency (Wirth, Glur, Gilli, & Anselmetti, 2013), erosion rate in Lake Verney, total solar irradiance (TSI) (Steinhilber, Beer, & Fröhlich, 2009), chironomid-based July air temperature in Lake Anterne reconstructed (Millet et al., 2009), Europe summer temperature anomaly (Büntgen et al., 2011), Great Aletsch glacier extension (Holzhauser, Magny, & Zumbuhl, 2005).

Phase A (5750~2620 cal. yr BP) was characterized by the dominating presence of *Pinus* (pine, probably Swiss stone pine *Pinus cembra*) sequences with constant presences of tall forb sequences (e.g. *Ranunculus* MOTU 2 and *Geranium*) and of some heath-shrub sequences such as Rhodoreae and Cupressaceae. Most meadow forbs were absent during Phase A, except some sporadic detections. These results suggest a pine forest community with a tall forb understory.

Phase B (2620~1030 cal. yr BP) began with a large drop of the frequency of pine sequences, accompanied by an increase of the frequency of sequences of meadow taxa, such as *Plantago* MOTU 1 and *Alchemilla* MOTUs 1 and 2, suggesting an abrupt opening of environment due to intensified human activities in this area. Despite the loss of pine trees, tall forb communities persisted. Pine DNA partially recovered during the later period of this phase, corresponding to the Early Middle Ages (1550-1030 cal. yr BP).

Phase C (1030 cal. yr BP ~ today) corresponded to the time interval from the beginning of the Medieval Climate Anomaly (MCA) to the modern period. It was characterized by a second wave of deforestation and by the decline of tall forb community, and correspondingly by a surge of meadow species DNA. Such composition remained stable through the MCA and the Little Ice Age (LIA, ca. 450~100 cal. yr BP, Badino et al., 2018).

Visualization of the principle component analysis (PCA) of plant communities was consistent with the grouping of samples (Figure 4-3).

a). In the PCA, the first two principle components accounted for 29.3% and 7.0% of the total variance, respectively. PC1 distinguished a *Pinus* (Pines) forest of Phase A with negative values, from the meadow communities of Phase C with positive values, character by grazing indicator taxa such as *Plantago* MOTUs and *Alchemilla* MOTUs (Figure 4-3b). Positive PC2 values were related to shrub communities characterized by taxa from the heath family (e.g. *Vaccinium*), and negative values to tall forb communities represented by e.g. *Rumex* and *Ranunculus* MOTU 2 (Figure 4-2, Figure 4-S3). PC2 is therefore able to distinguish between Phases B and C (Figure 4-3).

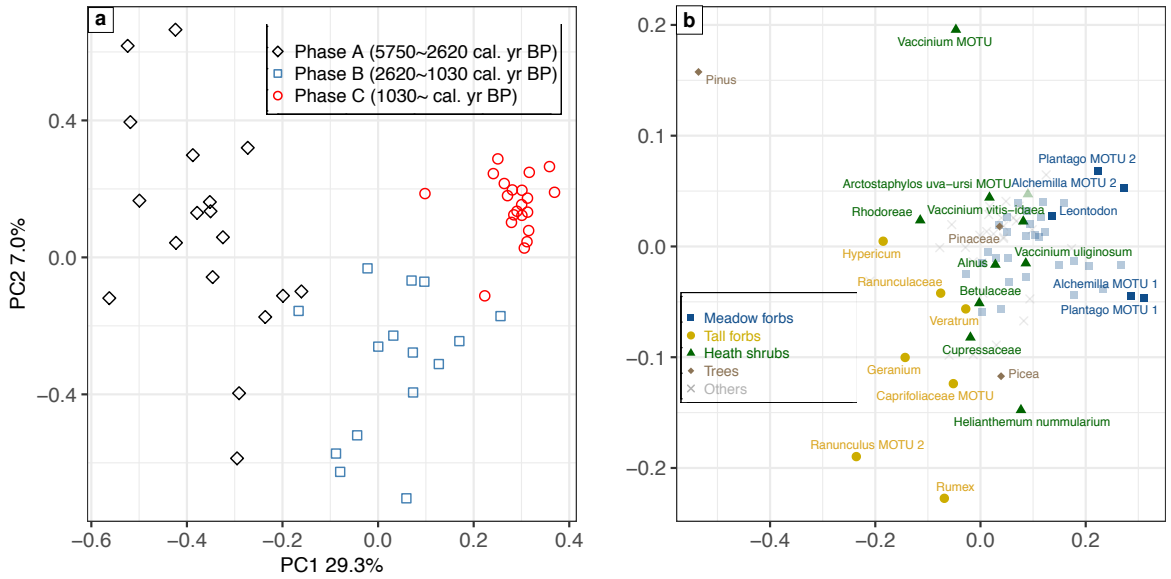


Figure 4-3: principle component analysis (PCA) of vegetation, showing sample distribution (a) and plant taxon distribution (right panel) on the first two PCA axes (b).

4.2.4 Occurrences of domestic animals

Within the whole Phase A, no domestic animal DNA was found in samples. During Phase B, *Ovis* (sheep) DNA was detected in four samples, scattered over the whole phase. Through Phase B, cattle DNA was detected in only one sample dated to 1810 cal. yr BP, but we found a few DNA reads of both sheep and cattle (read number below the detection threshold), in the oldest sample of this phase (2550 cal. yr BP). Cattle and sheep DNA was frequently detected in samples dated to the early stage of Phase C. The DNA of domestic animals tended to decline after the LIA, with no sheep DNA detected in the recent samples.

Spores of coprophilous fungi (*Sporormiella*) are a excellent proxy for the presence of herbivores (Baker, Bhagwat, & Willis, 2013; Doyen & Etienne, 2017; Etienne, Wilhelm, Sabatier, Reyss, & Arnaud, 2013), and even allow inferring their biomass (Gill et al., 2013), even though they do not allow the identification of herbivore species. The variation of spore accumulation rate through time was similar to that of the animal DNA, but in a smoother manner (Figure 4-2). Spore accumulation rate was generally low over the Phase A, despite a few minor peaks in some of the most recent samples. Spore accumulation rate

increased through Phase B from 20 to 100 no·cm^{-2}·yr^{-1}, reached the maximum values during Phase C before the LIA, and dropped thereafter to a level similar to those of the Phase B.

The accumulation rate of coprophilous fungus spores was significantly correlated with the number of positive cattle DNA replicates through the whole period (Spearman correlation: $r = 0.538$, $P = 2.74 \cdot 10^{-5}$), and, to a lesser extent, with the number of positive sheep DNA replicates (Spearman correlation: $r = 0.260$, $P = 0.0574$).

4.2.5 Erosion rate

Erosion rate was estimated from the terrigenous flux to the lake basin (Bajard, Poulenard, Sabatier, Etienne, et al., 2017). Through Phase A, erosion rate measured as terrigenous flux fluctuated at relatively low levels between 5.6 and 9.8 t·km^{-2}·yr^{-1}, with an average of 7.0 t·km^{-2}·yr^{-1}. Erosion started to rise at the beginning of Phase B and reached its maximum values during the Little Ice Age (26.9 t·km^{-2}·yr^{-1} in average, with a peak of 37.7 t·km^{-2}·yr^{-1} in 220 cal. yr BP). The erosion rate dropped after the LIA to ~10 t·km^{-2}·yr^{-1}, a level similar to the one observed at around 1900 cal. yr BP. Breakpoint analysis indicated a three-segment partition with two breakpoints at 2760 and 150 cal. yr BP, corresponding to the beginning of the rise of erosion rate and to the beginning of its drop, respectively. The first breakpoint at 2760 cal. yr BP is chronologically close to the boundary between vegetation Phases A and B at 2620 cal. yr BP.

4.2.6 Interactions among climate, livestock farming, plant community composition, and soil erosion

By plotting erosion rate against total solar irradiance [TSI, (Steinhilber et al., 2012, 2009)], we found that the response of soil erosion to climate variability was more evident after the first erosion rate breakpoint at 2760 cal. yr BP (Figure 4-4**a**). TSI is known to have an influence on past extreme precipitation in this Alpine area (Czymzik, Muscheler, & Brauer,

2016; Sabatier et al., 2017; Wilhelm et al., 2012; Wirth et al., 2013). We tested this hypothesis using linear regressions. Erosion rate decreased in periods with high solar irradiance after 2760 cal. yr BP (autoregressive model, $b_2 = -0.973 \pm 0.329$, $t_{106} = -5.79$, $P < 0.0001$), but not before ($b_1 = 0.0486 \pm 0.481$, $t_{156} = 0.198$, $P = 0.843$).

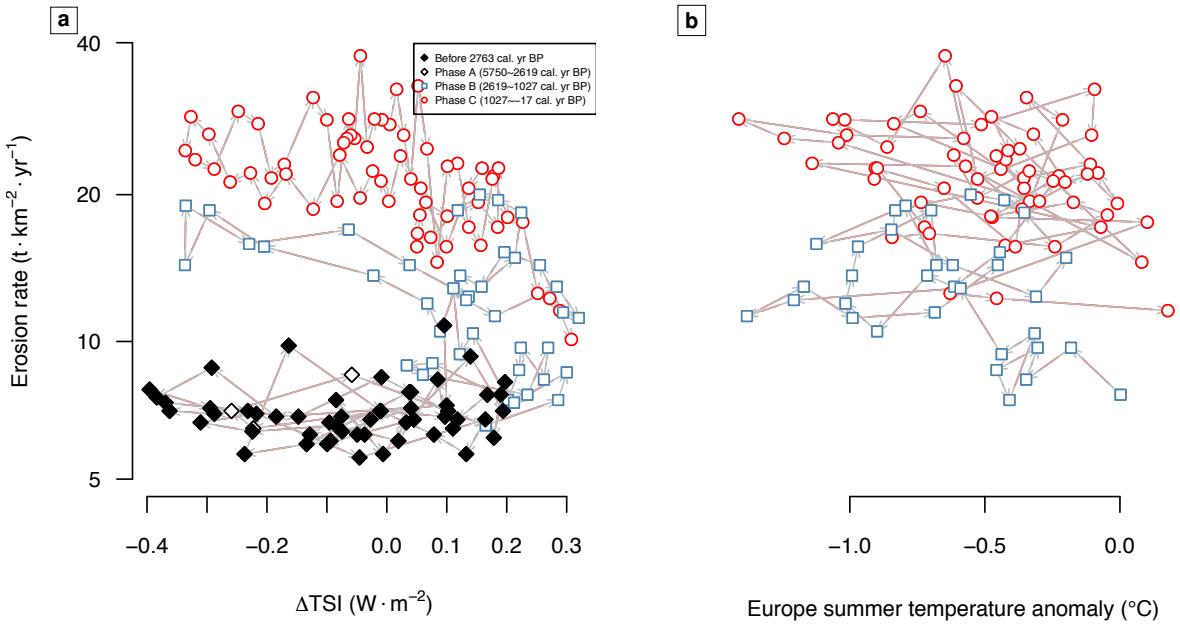


Figure 4-4: diagrams showing the relationship between Lake Verney erosion rate and climate variability. a) with total solar irradiance (TSI) (Steinhilber et al., 2009); b) with Europe summer temperature anomaly (Büntgen et al., 2011).

As there could be different paths through which climate variability and human activities could have impact soil erosion [e.g. by altering precipitation pattern or by modifying vegetation cover (Arnaud et al., 2016; Giguet-Covex et al., 2012)], we used piecewise Structural Equation Models (SEM) (Lefcheck, 2016) to assess the relationships among climate, livestock farming, plant communities and soil erosion. We hypothesized that erosion rate can be affected by vegetation changes and by climate, and that vegetation can be affected by livestock and climate (Figure 4-5). Our model assumed that erosion rate can be affected by livestock farming, vegetation changes and by climate, and that vegetation can be affected by livestock, reflected in the changes in vegetation PC1, and by climate, characterized by the changes in vegetation PC2. The piecewise SEM, based on autoregressive linear models, provided a good fit to the data (Fisher's C statistics $C_4 = 4.21$, χ^2 test $P = 0.378$). Erosion

rate was strongly influenced by the composition of vegetation (PC1 and PC2); furthermore, after 2760 cal. yr BP, TSI had a significant direct effect on erosion. Livestock pressure load, represented by the coprophilous fungi *Sporormiella*, was strongly related to vegetation PC1 (i.e. the component representing the gradient from pine forests and grazed meadow), but not to vegetation PC2. The direct effect of livestock pressure load on erosion rate was not significant. In fact, vegetation PC2 was significantly related to TSI after 2760 cal. yr BP. SEM provided an excellent fit to erosion rate ($R^2 = 0.778$).

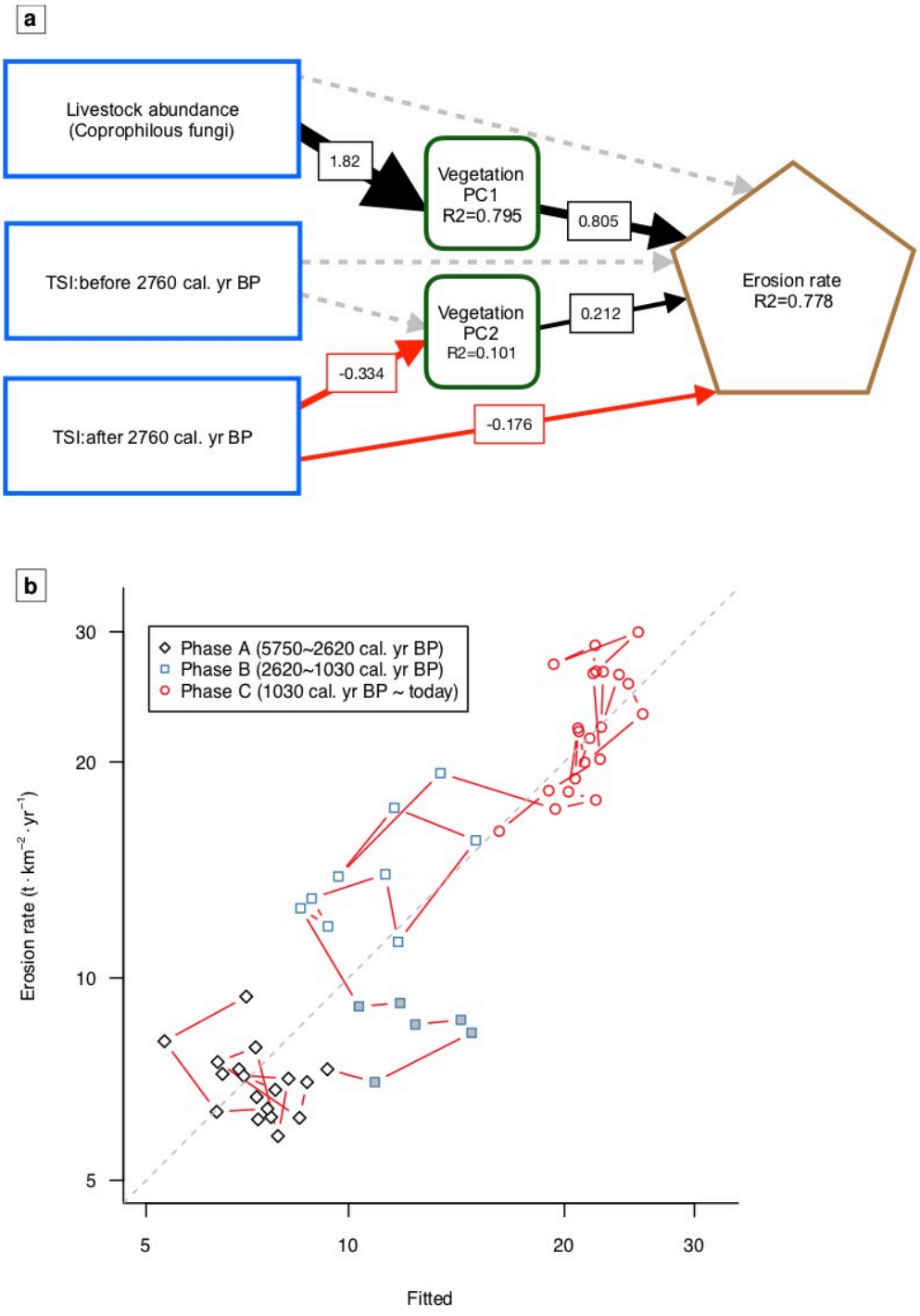


Figure 4-5: a) Structural equation model on the relationships among livestock abundance, total solar irradiance (TSI), vegetation (PC1 and PC2), and erosion rate. Solid arrows represent positive (black) or negative (red) significant relationship, with numbers in front of the arrow line indicating the relative magnitude of relation. Dash arrows in gray represent none-significant relationships. b) Scatter plot of erosion rate versus values fitted by the SEM.

4.3 Discussion

The combination between sedDNA and long-term paleoenvironmental information allowed to reconstruct the long-term development of vegetation in an alpine environment subjected to human influences over the last six thousand years, and to understand the complex impact of human activities on ecosystem functioning. The metabarcoding approach (Taberlet et al., 2018) provided an excellent account of catchment-scale vegetation change at high taxonomic- and temporal resolutions, and covering the major life forms of plants, from trees to shrubs and forbs. Such accurate community data allowed to disentangle the complex interactions among climate change, human activities, vegetation, and soil erosion. Our results highlight the complex interplay of the paths through which climate change and human activities modified vegetation and geomorphology of alpine environments during the second half of the Holocene. We observed not only an increase of soil erosion associated to the loss of forests and livestock farming, but also that soil erosion processes became more sensitive to climate variability after the major vegetation change, and we show the value of sedimentary DNA for reconstructing the multi-faceted dynamics of past environment changes, resolving complex long-term human-environment interactions (Ficetola, Poulenard, et al., 2018; Giguet-Covex et al., 2014; Madeja, Wacnik, Wypasek, Chandran, & Stankiewicz, 2010).

Human activities at high altitudes (mostly livestock farming in the past millennia) often act as promoters of soil erosion by modifying vegetation cover (Brisset et al., 2017; Giguet-Covex et al., 2014), and by destabilizing soils (Dapples et al., 2002). Our analyses confirmed the human human role for both the shift of vegetation type and the change of erosion pattern. First, the largest changes in plant community composition, characterized by the retreat of pine forest and the expansion of grazed meadows (vegetation PC1, see Figure 4-3), were driven by livestock farming (Figure 4-3, Figure 4-5). This result is consistent with previous observations in similar environments (e.g. Pansu, Giguet-Covex, et al.,

2015), and may represent a common trend of human-induced deforestation in the Alps during the late Holocene (Pini et al., 2017). Second, such changes had a strong effect on erosion rate (Figure 4-5). Indeed, deforestation is often associated with increased soil erosion, when the protection provided by plant cover has dropped considerably (Arnaud et al., 2016; Bajard, Poulenard, Sabatier, Etienne, et al., 2017; Giguet-Covex et al., 2014). This result is in accordance with the widely observed threshold effect of vegetation cover over soil erosion (i.e. the existence of a threshold vegetation cover below which erosion increases markedly, Evans, 1998), as deforestation and grazing inevitably leave more bare soils.

Flood induced by extreme precipitation event is a major cause of soil erosion (Giguet-Covex et al., 2011, 2012; Vannière et al., 2013); therefore climate change, which has a strong influence on flood dynamics, is also an important driver on soil erosion (Arnaud et al., 2016; Giguet-Covex et al., 2012). TSI is a proxy of climate forcing associated with extreme precipitation, and consequently with flood in Western Alps (Czymzik et al., 2016; Sabatier et al., 2017; Wilhelm et al., 2012; Wirth et al., 2013), because lower TSI indicates that less heat from the sun arrives at the earth's surface, causing a colder climate (Gray et al., 2010), which subsequently translates to more extreme precipitation events in the Alps (Amann, Szidat, & Grosjean, 2015; Glur et al., 2013). The link between long-term climatic variation and TSI is confirmed by the increase of erosion in periods with low TSI (Figures). However, the negative relationship between the TSI and erosion was only evident after a breakpoint occurring 2760 cal. yr BP, which suggested a major change in the regime of local geomorphology. The date of and that of this erosion breakpoint (2760 cal. yr BP) matches well the first major vegetation change (2620 cal. yr BP), suggesting an indirect impact of human-induced landscape change upon soil erosion via affecting its susceptibility to extreme precipitation events and thus to climate variability. Despite the impact of climate and land-use change on erosion are well known, disentangling the complex interactions between these effects has been challenging. Previous studies often treated the two forcing factors separately, focusing on their relative importance through the past millennia (e.g. Vannière et al., 2013). For the first time, our approach allowed to formally test the complex interactions among climate, land-use and soil degradation processes,

thanks to the highly localized, high-resolution vegetation signals provided by sedimentary DNA. On the other hand, heavy livestock grazing compacts and, destabilizes bare soil, reduces water infiltration (Evans, 1998; Trimble & Mendel, 1995), therefore it can also directly increase soil sensitivity to extreme precipitation events. However, we found no significant direct effect of livestock abundance on erosion rate (Figure 4-5). This might be owed to the lack of statistical power rather than the absence of such effect, since the limited sample size (50) might not have been able to allow separating such direct effect from that which was mediated by vegetation change. Alternatively, the impact of overgrazing may also be perceived through an increasing sensitivity of soil erosion to climate change along with rising livestock abundance. Again, our small dataset would not allow testing this hypothesis with many variables in play. Nevertheless, future studies may elucidate this mechanism with more data.

SedDNA provided accurate information on environmental changes occurring at the landscape (i.e. catchment) scale, still they can allow to understand the past climate-vegetation-human-soil development/erosion relationships upon a much broader scale. The erosion pattern of the Lake Verney catchment is fairly similar to that of the Northern Alps stacked flood frequency (see Figure 4-2, Wirth et al., 2013) over the last 5800 years, and similar histories of deforestation and livestock farming were observed in sites along the Alps range (Dietre et al., 2014; Giguet-Covex et al., 2014; Röpke et al., 2011), suggesting that human impact through vegetation and or livestock farming changes have probably too much neglected in past flood chronicles (Brisset et al., 2017).

Therefore, our results would help understand the geomorphology in the Alps during the mid- and late Holocene. Most importantly, for the first time our results integrate different aspects of those complex relationships previously only investigated in separated projects (Brisset et al., 2017). Furthermore, since soil degradation is a crucial issue of ecosystem functioning since the first large-scale land-use changes (Vanwalleghem et al., 2017), long-term observational data is a key component of our knowledge to guide a sustainable land-use management in a global change context. The present study thus also provides valuable

information for future management to preserve ecosystem functioning and services in mountainous areas.

4.4 Material and methods

4.4.1 Study site

Lake Verney (2088 m a.s.l.) is located in the Western Alps, on the Italian side of the Petit Saint Bernard Pass (2188 m a.s.l.) (Figure 4-1) between Séez in the Tarentaise valley (France) and La Thuile in the Aoste valley (Italy). The lake is fed by three main streams that flow from the south of the lake. Its catchment covers an area of 3.5 km² and rises up to 2900 m a.s.l. The surface of the lake is 0.2 km² and a maximum depth of 21 m. The coring took place in a shallower part of the lake at ca. 15 m (Figure 4-1b).

4.4.2 SedDNA analyses

We sampled 55 slices (2-cm thickness) from the sediment core. Sampling details (depths, thicknesses, weights) are presented in Table S1. In order to avoid contamination, the edges of slices were removed with sterile disposable scalpels. DNA extraction targeted extracellular DNA (Pansu, Giguet-Covex, et al., 2015). For each sediment slice, we mixed about 15 g of sediment with 15 ml of saturated phosphate buffer (Na2HPO4; 0.12 M; pH ≈ 8) during 15 min. We then centrifuged 15 ml of the mixture (10min at 10000 g). 12 ml of the resulting supernatant were transferred to Amicon ® Ultra-15 10K Centrifugal Filter Devices (Millipore). We then centrifuged (20 min at 4000 g) the devices for ultrafiltration and concentration of DNA. We kept 400 µl of the resulting concentrate for further extraction with the NucleoSpin® Soil kit (Macherey-Nagel, Düren, Germany). The cell lysis step was skipped, the remaining steps followed manufacturer's instructions as described in (Taberlet, Prud'Homme, et al., 2012). Four extraction controls were performed.

We targeted mammal and vascular plant DNA for amplification. Mammal DNA was amplified with a primer pair targeting a 60 to 84bp fragment of the mitochondrial 16S gene

(refferred to as P007 in Giguet-Covex et al., 2014; which corresponds to Mamm02 of Taberlet, Bonin, Zinger, & Coissac, 2018). For vascular plant DNA, we used the *g*-*h* primers (Taberlet et al., 2007, coresponding to Sper01 in Taberlet et al., 2018) targeting the P6 loop region of chloroplast *trn*L intron, a fragment whose length varies from 10 to 220 bp. We added 8 bp tags (with at least 5 differences between them) to the 5' end of primers in order to assign sequence reads to the relevant sample (Binladen et al., 2007; Valentini, Miquel, et al., 2009). DNA amplifications were carried out in a final volume of 20 µL containing 2 µL of diluted DNA extract, 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer and 4.8 µg of bovine serum albumin (BSA, Roche Diagnostic). To limit the amplification of human DNA when using the Mamm02 primers, we also added human-specific, blocking oligonucleotide (MamP007_B_Hum1). After 10 min at 95°C, we performed 45 PCR cycles: 30 s at 95°C, 30 s at 50°C, and 1 min at 72°C with a final elongation step (7 min at 72°C). We also amplified two PCR negative controls, containing PCR mix but no DNA template and two PCR positive controls, each containing a mix of 0.18 ng of DNA extracted from taxa absent in Europe [the marsupial *Dideplhis marsupialis* for mammal DNA and *Mouriri* sp., a tropical plant from the South America, belonging to the family of Melastomataceae] (Parducci, Bennett, Ficetola, Alsos, Suyama, Wood, & Pedersen, 2017). All samples and controls were amplified in 12 replicated PCRs (Ficetola et al., 2015). We purified and mixed the final PCR products before sequencing, which was performed by 2 x 125 bp pair-end sequencing on Illumina HiSeq 2500 platform.

Paired-end sequences were aligned and merged, and identical sequences were merged together with their information of original samples being kept. Merged sequences with alignment score lower than 40 were discarded. To ensure the quality of the SedDNA results, we filtered the sequences using the OBITOOLS software package (Boyer et al., 2016). We retained only those sequences, which contained both primers with no more than two mismatches in each as well as the exact sample tag sequence. All sequences shorter than 10 bp or containing ambiguous nucleotides, as well as sequences occurring only once in the dataset were discarded. To filter out potential amplification and sequencing errors, we built networks of sequences by linking any sequences differing by a single nucleotide

substitution or indel using the *obiclean* command. According to it corresponding network, each sequence fell into one of three categories: 'heads', which were only connected to sequences with a lower count, 'singletons', which were not connected to any other sequence and 'internals', which did not fall into either of the categories above. Sequences falling more often into 'internal' than 'head' or 'singleton' were then considered to be likely erroneous sequences and discarded.

Each remaining unique sequence formed a molecular operational taxonomic unit (MOTU). Taxonomic labels were assigned to each MOTU by aligning to sequences in a database generated by *in silico* PCR (Ficetola et al., 2010) on the EMBL release 136 nucleotide database. For plants, the taxonomic alignment was firstly performed on the global database including all taxa in the EMBL database and then on a database with only alpine species (Pansu, Giguet-Covex, et al., 2015). Inconsistences between the two resulting assigned datasets were resolved based on knowledge about the present local vegetation. For mammals, we kept only the MOTUs with a match higher than 97% to a mammal genus and with more than 5 reads in a sample (Sabatier et al., 2017). Given the higher abundance of plant DNA, for vascular plants we kept only those MOTUs with more 1000 reads across all samples, assigned to the family or better taxonomic levels (genus, species…), and with a 100% match to a taxon. A MOTU in a particular PCR replicate was considered to be positive if it had >10 reads, given that we consistently obtained much more plant than mammal reads. MOTUs with >3 positive replicates across all the controls were removed from the dataset. MOTUs assigned to exotic taxa were obvious contaminants and also discarded. The remaining MOTUs were kept for statistical analyses.

4.4.3 Estimating the erosion

As the coring was not realized in the deepest part of the lake (21 m), we cannot consider all the lake filling and whole catchment area to estimate the erosion. Estimation of the erosion in the catchment was calculated according a simplified sediment filling model following the method described in Bajard, Poulenard, Sabatier, Etienne, et al., (2017), considering the

surface of the lake that collects sediments around the coring point (dashed blue line in Figure 4-1b). This surface is estimated as a circle of 105 m of radius, i.e. 0.035 km^2. The area on the shore of the lake was excluded of this surface as it displayed sub-aquatic rocks and therefore less sediment. The shape of the lake filling was assimilated to a half ellipsoid volume, centered on the circle of radius 105 m at the surface of the lake. The volume (V) of each layer was calculated by subtracting two half ellipsoids of minor axe equal to the difference of sampling depth, based on the 2.4 m of the sediment core. The minor axis corresponded to the depth of water at the coring point (i.e. 15 m) plus the depth of sediment, based on the sampling depth of dry density measures. The dry density (DD) of the sediment was calculated in continue every 1 cm from the dry weight of a known volume of sediment along the 2.4 m of sequence to get the total mass of sediment at each depth. The average dry density was 0.75 $g.cm^{-3}$. Only the terrigenous component of the sediment was considered using the final LOI residue, i.e., Non-Carbonate Ignition Residue (NCIR) after the 950°C ignition phase, to estimate the erosion (E). The mass was integrated according to the time (T) that was covered by the layer of sediment as deduced from the age-depth model. Following this approach, the OM and carbonate fractions were neglected in the sediment budget because these factors could be produced and precipitated within the lake. This yields to an underestimation of the accumulation rate of sediment. Then, this accumulation rate of sediment was divided by the potential surface of erosion (S), which corresponds to the surface of the catchment expected to bring sediments in the area of lake used to estimate the filling (dashed yellow line in Figure 4-1b). Areas of scree were excluded of this surface which was finally set as 0.39 km^2.

Finally, the erosion equation could be written as:

$$E\ [t/km^2/yr] = \frac{DD \times V \times NCIR}{T \times S}$$

DD [g/cm^3], V [m^3], NCIR %, T [yr], S [km^2]

4.4.4 Coprophilous fungi spores

Following the same sampling that of sedDNA, 51 subsamples of approximately 1 cm^3 were prepared for non-pollen palynomorphs (NPPs) analysis, using the standard procedure described by Faegri and Iversen (1989). *Lycopodium clavatum* tablets were added to each subsample (Stockmarr, 1971) in order to calculate the concentration of NPPs. *Sporormiella* sp. were counted following the procedure described by Etienne and Jouffroy-Bapicot (2014) and are expressed as accumulation rates ($no \cdot cm^{-2} \cdot yr^{-1}$). Their ecological interpretation in term of agro-pastoral activities and OM origin follows the work of Doyen and Etienne (2016). Four samples, dated to 4220, 4380, 4530 and 6340 cal. yr BP, respectively, were missing during the measuring process; therefore their accumulation rates were imputed from the average accumulation rate before 2760 cal. yr BP, i.e. the first breakpoint of erosion rate.

4.4.5 Statistical analyses

We used a series of statistical analyses to characterize the development of plant communities, based on the taxon-sample matrix of number of positive replicates, which is considered as a measure of taxon occupancy (Ficetola et al., 2016) and even a measure of DNA abundance (Ficetola et al., 2015; Furlan et al., 2016). The matrix was Hellinger-transformed before analyses (Legendre & Legendre, 2012). First, we used multivariate regression trees (MRT) to test if the communities could be divided into several temporally consecutive clusters with similar compositions. MRT is a statistical technique that allows hierarchically clustering multivariate objects by repeated splitting of the data, based on simple rules applied to environment variables (De'ath, 2002), i.e. age for the community time sequence in our case (Legendre & Gauthier, 2014). The analysis of similarity (ANOSIM) was performed in order to assess the significance of the differences between each pair of the resulting clusters. Then we used principle component analysis (PCA) to visualize the distribution of samples, as well as the clusters, and to summarize the variation of plant communities along a limited number of axes. The first two principle components (vegetation PC1 and PC2) represented 29.3% and 7.0% of the total variance, respectively.

As proxies of livestock farming, we used the numbers of positive PCR replicates of both cattle and sheep sequences. As a complementary measure to mammal DNA (Ficetola, Poulenard, et al., 2018), we also used the abundance of the spores of coprophilous fungus *Sporormiella* (Bajard, Poulenard, Sabatier, Etienne, et al., 2017) (log-transformed accumulation rate), which had been demonstrated to be able to reflect the presence and the even density of large herbivores (Baker et al., 2013; Doyen & Etienne, 2017; Etienne et al., 2013; Gill et al., 2013). To trace the variability of potential climatic factors (e.g. temperature, precipitation) affecting the local plant communities and erosion, we adopted the total solar irradiance (TSI) time series as proxy (Steinhilber et al., 2009). In fact, very few long-term records with enough resolution were available to represent climate change in this region. As the forcing of the TSI is a major factor of long-term climatic variability, it was used as a proxy of climate to identify warm and cold periods (Gray et al., 2010).

We used terrigenous flux (Bajard, Poulenard, Sabatier, Etienne, et al., 2017) as a measure of soil erosion through time; terrigenous flux was log-transformed to improve normality. We dated and detected potential breakpoints in the erosion time sequence by a dynamic programming algorithm that searches for structural changes in a time series and dates the breakpoints (Zeileis, 2003). We then tested the hypothesis that the erosion rate was more susceptible to climate variability after the first (older) breakpoint then before, as it can be viewed by plotting the erosion rate against the TSI (Figure 4-4). To do so, we performed a generalized least-squares (GLS) regression, taking the TSI as explanatory variable and erosion rate as dependent, conditioned on a factor indicating the age before or after the first breakpoint. Potential temporal autocorrelation in residuals were modeled with a first-order autoregressive term (Jones, 1993).

On the other hand, the effect of the same climate variability can be different or even reversed in different time scales (Mills et al., 2016). To confirm the unequal sampling time intervals of erosion rate did not confound the results, the same linear regression was also

performed on a dataset with 200-year running averages of both TSI and erosion rate. The results were identical with both pairs of tests.

Based on the results of the above analyses, we established a causal model to link climate variability and livestock farming to plant community development and soil erosion. The model was based on generalized least-squares (GLS) regressions, with residuals modeled by a first-order autoregressive term to deal with potential temporal autocorrelations (Figure 4-5). We used piecewise structural equation modeling [piecewise SEM, (Lefcheck, 2016)] to confirm that the proposed model is appropriate to represent the data. Piecewise structural equation modeling (piecewise SEM) is a statistical tool for confirmatory path analysis that tests multivariate causal relationships involving only observed variables (Lefcheck, 2016; Shipley, 2000b, 2009). The goodness-of-fit of the piecewise SEM was evaluated by testing whether the potentially missing causal paths not specified by the model were significant. This was done by calculating the Fisher's *C* statistic, which measures the significance of the potentially missing paths, and by comparing this statistic to a chi-square distribution with $2k$ degrees of freedom, k being the size of the minimum set of such missing paths (Lefcheck, 2016; Shipley, 2009).

4.5 Acknowledgements

GFF has received funding from the European Research Council under the European Community Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). The Laboratoire d'Écologie Alpine is part of Labex OSUG@2020 (ANR10 LABX56).

4.6 Data Accessibility

- Filtered eDNA (mammal and plant) sequencing data, plant reference database and geochemical and sedimentological data are available in the Dryad Digital Repository doi:
-

4.7 Author Contributions

G.F.F., F.A. and P.T. conceived the study. P.S., F.A., M.B., and G.F.F. performed the fieldwork. G.F.F., P.T., D.R., P.S., J.P., D.E. and J.L.B. performed the laboratory work. W.C. performed the analyses with subsequent contributions by G.F.F., E.M., P.S., J.P. and M.B. W.C. wrote the first draft of the manuscript, with subsequent contribution by all the authors.

4.8 Reference

Alsos, I. G., Sjögren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., … Pedersen, M. W. (2016). Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene, 26(4), 627–642. http://doi.org/10.1177/0959683615612563

Amann, B., Szidat, S., & Grosjean, M. (2015). A millennial-long record of warm season precipitation and flood frequency for the North-western Alps inferred from varved lake sediments: Implications for the future. Quaternary Science Reviews, 115(January 2018), 89–100. http://doi.org/10.1016/j.quascirev.2015.03.002

Arnaud, F., Poulenard, J., Giguet-Covex, C., Wilhelm, B., Révillon, S., Jenny, J. P., … Sabatier, P. (2016). Erosion under climate and human pressures : An alpine lake sediment perspective. Quaternary Science Reviews, 152, 1–18. http://doi.org/10.1016/j.quascirev.2016.09.018

Badino, F., Ravazzi, C., Vallè, F., Pini, R., Aceti, A., Brunetti, M., … Orombelli, G. (2018). 8800 years of high-altitude vegetation and climate history at the Rutor Glacier

forefield, Italian Alps. Evidence of middle Holocene timberline rise and glacier contraction. Quaternary Science Reviews, 185, 41–68. http://doi.org/10.1016/j.quascirev.2018.01.022

Bajard, M., Poulenard, J., Sabatier, P., Develle, A.-L. L., Giguet-Covex, C., Jacob, J., … Arnaud, F. (2017). Progressive and regressive soil evolution phases in the Anthropocene. Catena, 150, 39–52. http://doi.org/10.1016/j.catena.2016.11.001

Bajard, M., Poulenard, J., Sabatier, P., Etienne, D., Ficetola, F., Chen, W., … Arnaud, F. (2017). Long-term changes in alpine pedogenetic processes: effect of millennial agro-pastoralism activities (French-Italian Alps). Geoderma, 306(June), 217–236. http://doi.org/10.1016/j.geoderma.2017.07.005

Bajard, M., Sabatier, P., David, F., Develle, A. L., Reyss, J. L., Fanget, B., … Arnaud, F. (2015). Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. Holocene, 26(3), 350–364. http://doi.org/10.1177/0959683615609750

Baker, A. G., Bhagwat, S. A., & Willis, K. J. (2013). Do dung fungal spores make a good proxy for past distribution of large herbivores? Quaternary Science Reviews, 62, 21–31. http://doi.org/10.1016/j.quascirev.2012.11.018

Bärring, L., Jönsson, P., Mattsson, J. O., & Åhman, R. (2003). Wind erosion on arable land in Scania, Sweden and the relation to the wind climate - A review. Catena, 52(3–4), 173–190. http://doi.org/10.1016/S0341-8162(03)00013-4

Barrios, E. (2007). Soil biota, ecosystem services and land productivity. Ecological Economics, 64(2), 269–285. http://doi.org/10.1016/j.ecolecon.2007.03.004

Berendse, F., van Ruijven, J., Jongejans, E., & Keesstra, S. (2015). Loss of plant species diversity reduces soil erosion resistance. Ecosystems, 18(5), 881–888. http://doi.org/10.1007/s10021-015-9869-6

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PloS One, 2(2), e197.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. http://doi.org/10.1111/1755-0998.12428

Brisset, E., Guiter, F., Miramont, C., Troussier, T., Sabatier, P., Poher, Y., … Anthony, E. J. (2017). The overlooked human influence in historic and prehistoric floods in the European Alps. Geology, 45(4), 347–350. http://doi.org/10.1130/G38498.1

Büntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., … Esper, J. (2011). 2500 years of European climate variability and human susceptibility. Science, 331(6017), 578–582. http://doi.org/10.1126/science.1197175

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., … Naeem, S. (2012). Biodiversity loss and its impact on humanity. Nature, 486(7401), 59–67. http://doi.org/10.1038/nature11148

Czymzik, M., Muscheler, R., & Brauer, A. (2016). Solar modulation of flood frequency in central Europe during spring and summer on interannual to multi-centennial timescales. Climate of the Past, 12(3), 799–805. http://doi.org/10.5194/cp-12-799-2016

Dapples, F., Lotter, A. F., Van Leeuwen, J. F. N., Van Der Knaap, W. O., Dimitriadis, S., & Oswald, D. (2002). Paleolimnological evidence for increased landslide activity due to forest clearing and land-use since 3600 cal bp in the western swiss alps. Journal of Paleolimnology, 27(2), 239–248. http://doi.org/10.1023/A:1014215501407

De'ath, G. (2002). Multivariate Regression Tree: A New Technique for Modeling Species–Environment Relationships. Ecology, 83(4), 1105–1117. http://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2

Dearing, J. A. (2008). Landscape change and resilience theory: A palaeoenvironmental assessment from Yunnan, SW China. Holocene, 18(1), 117–127. http://doi.org/10.1177/0959683607085601

Dietre, B., Walser, C., Lambers, K., Reitmaier, T., Hajdas, I., & Haas, J. N. (2014). Palaeoecological evidence for Mesolithic to Medieval climatic change and anthropogenic impact on the Alpine flora and vegetation of the Silvretta Massif (Switzerland/Austria). Quaternary International, 353, 3–16. http://doi.org/10.1016/j.quaint.2014.05.001

Dotterweich, M. (2013). The history of human-induced soil erosion: Geomorphic legacies, early descriptions and research, and the development of soil conservation-A global synopsis. Geomorphology, 201, 1–34. http://doi.org/10.1016/j.geomorph.2013.07.021

Doyen, E., & Etienne, D. (2017). Ecological and human land-use indicator value of fungal spore morphotypes and assemblages. Vegetation History and Archaeobotany, 26(4), 357–367. http://doi.org/10.1007/s00334-016-0599-2

Edwards, M. E., Alsos, I. G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., … Taberlet, P. (2018). Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. The Holocene, 095968361879809. http://doi.org/10.1177/0959683618798095

Etienne, D., Wilhelm, B., Sabatier, P., Reyss, J. L., & Arnaud, F. (2013). Influence of sample location and livestock numbers on Sporormiella concentrations and accumulation rates in surface sediments of Lake Allos, French Alps. Journal of Paleolimnology, 49(2), 117–127. http://doi.org/10.1007/s10933-012-9646-x

Evans, R. (1998). The erosional impacts of grazing animals. Progress in Physical Geography, 22(2), 251–268. http://doi.org/10.1177/030913339802200206

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 1. http://doi.org/10.1186/1471-2164-11-434

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., … Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543–556. http://doi.org/10.1111/1755-0998.12338

Ficetola, G. F., Poulenard, J., Sabatier, D., Messager, E., Gielly, L., Leloup, A., … Arnaud, F. (2018). DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Science Advances. Science Advances. http://doi.org/10.1126/sciadv.aar4292

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources, 16(3), 604–607. http://doi.org/10.1111/1755-0998.12508

Fitter, A. H., Gilligan, C. A., Hollingworth, K., Kleczkowski, A., Twyman, R. M., Pitchford, J. W., & of the NERC Soil Biodiversity Programme, M. (2005). Biodiversity and ecosystem function in soil. Functional Ecology, 19(3), 369–377. http://doi.org/10.1111/j.1365-2435.2005.00969.x

Folke, C. (2006). Resilience: The emergence of a perspective for social-ecological systems analyses. Global Environmental Change, 16(3), 253–267. http://doi.org/10.1016/j.gloenvcha.2006.04.002

Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. Molecular Ecology Resources, 16(3), 641–654. http://doi.org/10.1111/1755-0998.12483

Giguet-Covex, C., Arnaud, F., Enters, D., Poulenard, J., Millet, L., Francus, P., … Delannoy, J. J. (2012). Frequency and intensity of high-altitude floods over the last 3.5ka in northwestern French Alps (Lake Anterne). Quaternary Research, 77(1), 12–22. http://doi.org/10.1016/j.yqres.2011.11.003

Giguet-Covex, C., Arnaud, F., Poulenard, J., Disnar, J. R., Delhon, C., Francus, P., … Delannoy, J. J. (2011). Changes in erosion patterns during the holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (lake anterne, 2063 m a.s.l., NW french alps): The role of climate and human activities. Holocene, 21(4), 651–665. http://doi.org/10.1177/0959683610391320

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., … Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications, 5, 3211. http://doi.org/10.1038/ncomms4211

Gill, J. L., Mclauchlan, K. K., Skibbe, A. M., Goring, S., Zirbel, C. R., & Williams, J. W. (2013). Linking abundances of the dung fungus sporormiella to the density of bison: Implications for assessing grazing by megaherbivores in palaeorecords. Journal of Ecology, 101(5), 1125–1136. http://doi.org/10.1111/1365-2745.12130

Glur, L., Wirth, S. B., Büntgen, U., Gilli, A., Haug, G. H., Schär, C., … Anselmetti, F. S. (2013). Frequent floods in the European Alps coincide with cooler periods of the past 2500 years. Scientific Reports, 3, 1–5. http://doi.org/10.1038/srep02770

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., … White, W. (2010). Solar influence on climate. Reviews of Geophysics, 48(2009), RG4001. http://doi.org/10.1029/2009RG000282.1.INTRODUCTION

Herman, F., Seward, D., Valla, P. G., Carter, A., Kohn, B., Willett, S. D., & Ehlers, T. A. (2013). Worldwide acceleration of mountain erosion under a cooling climate. Nature, 504(7480), 423–426. http://doi.org/10.1038/nature12877

Jones, R. H. (1993). Longitudinal data with serial correlation: a state-space approach. Chapman and Hall/CRC.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods in Ecology and Evolution, 7(5), 573–579. http://doi.org/10.1111/2041-210X.12512

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space − time analysis of community composition data. Proceedings of the Royal Society of London B: Biological Sciences, 281(1778), 20132728. http://doi.org/http://dx.doi.org/10.1098/rspb.2013.2728

Legendre, P., & Legendre, L. F. J. (2012). Numerical ecology (3rd ed.). Amsterdam: Elsevier.

Loreau, M., & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: A synthesis of underlying mechanisms. Ecology Letters, 16(SUPPL.1), 106–115. http://doi.org/10.1111/ele.12073

Ludwig, J., Wilcox, B. P., Breshears, D. D., Tongway, D. J., & Imeson, A. C. (2005). Vegetation patches and runoff- erosion as interacting ecohydrological processes in semiaris landscapes. Ecology, 86(2), 288–297. http://doi.org/10.1890/03-0569

Madeja, J., Wacnik, A., Wypasek, E., Chandran, A., & Stankiewicz, E. (2010). Integrated palynological and molecular analyses of late Holocene deposits from Lake Miłkowskie (NE Poland): Verification of local human impact on environment. Quaternary International, 220(1–2), 147–152. http://doi.org/10.1016/j.quaint.2009.09.008

Martin, C., Pohl, M., Alewell, C., Körner, C., & Rixen, C. (2010). Interrill erosion at disturbed alpine sites: Effects of plant functional diversity and vegetation cover. Basic and Applied Ecology, 11(7), 619–626. http://doi.org/10.1016/j.baae.2010.04.006

Metzger, M. J., Rounsevell, M. D. A., Acosta-Michlik, L., Leemans, R., & Schröter, D. (2006). The vulnerability of ecosystem services to land use change. Agriculture, Ecosystems and Environment, 114(1), 69–85. http://doi.org/10.1016/j.agee.2005.11.025

Mills, K., Schillereff, D., Saulnier-Talbot, É., Gell, P., Anderson, N. J., Arnaud, F., … Mcgowan, S. (2016). Deciphering long ‐ term records of natural variability and human impact as recorded in lake sediments : a palaeolimnological puzzle. Wiley Interdisciplinary Reviews, 4(2), e1195. http://doi.org/10.1002/wat2.1195

Miras, Y., Millet, L., Guiter, F., Ponel, P., De Beaulieu, J. L., & Tozlar, T. (2006). Dynamique des Ecosystèmes et Impact de l'homme dans le Secteur du Col du Petit Saint Bernard au Cours de l'Holocène. In Alpis Graia: Archéologie sans frontieres au col du Petit-Saint-Bernard (pp. 31–50).

Mohammad, A. G., & Adam, M. A. (2010). The impact of vegetative cover type on runoff and soil erosion under different land uses. Catena, 81(2), 97–103. http://doi.org/10.1016/j.catena.2010.01.008

Nearing, M. a, Pruski, F. F., & O'Neal, M. R. (2004). Expected climate change impacts on soil erosion rates: A review. Journal of Soil and Water Conservation, 59(1), 43–50. Retrieved from http://www.jswconline.org/content/59/1/43.abstract

Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J. M., Gielly, L., Cordonnier, T., … C??cillon, L. (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA. Soil Biology and Biochemistry, 83, 100–105. http://doi.org/10.1016/j.soilbio.2015.01.004

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. Molecular Ecology, 24(7), 1485–1498. http://doi.org/10.1111/mec.13136

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., … Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924–942. http://doi.org/10.1111/NPH.14470

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist.

Peterson, G., Allen, C. R., & Holling, C. S. (1998). Ecological Resilience , Biodiversity , and Scale. Ecosystems, 1, 6–18. http://doi.org/10.1007/s100219900002

Pini, R., Ravazzi, C., Raiteri, L., Guerreschi, A., Castellano, L., & Comolli, R. (2017). From pristine forests to high-altitude pastures: an ecological approach to prehistoric human impact on vegetation and landscapes in the western Italian Alps. Journal of Ecology. http://doi.org/10.1111/1365-2745.12767

Rey, P.-J., & Moulin, B. (2011). Premières occupations de la montagne alpine sur les versants du Petit Saint Bernard (programme Alpis Graia). Principaux résultats des prospections et sondages archéologiques sur le versant valdôtain (années 2005-2006). Bulletin d'études Préhistoriques et Archéologiques Alpines, (XXII), p--9.

Röpke, A., Stobbe, A., Oeggl, K., Kalis, A. J., & Tinner, W. (2011). Late-holocene land-use history and environmental changes at the high altitudes of st antönien (switzerland, northern alps): Combined evidence from pollen, soil and tree-ring analyses. Holocene, 21(3), 485–498. http://doi.org/10.1177/0959683610385727

Sabatier, P., Dezileau, L., Colin, C., Briqueu, L., Bouchette, F., Martinez, P., … Von Grafenstein, U. (2012). 7000 years of paleostorm activity in the NW Mediterranean Sea in response to Holocene climate events. Quaternary Research, 77(1), 1–11. http://doi.org/10.1016/j.yqres.2011.09.002

Sabatier, P., Poulenard, J., Fanget, B., Reyss, J.-L., Develle, A.-L., Wilhelm, B., … Arnaud, F. (2014). Long-term relationships among pesticide applications, mobility, and soil erosion in a vineyard watershed. Proceedings of the National Academy of Sciences, 111(44), 15647–15652. http://doi.org/10.1073/pnas.1411512111

Sabatier, P., Wilhelm, B., Ficetola, G. F., Moiroux, F., Poulenard, J., Develle, A.-L., … Arnaud, F. (2017). 6-kyr record of flood frequency and intensity in the western Mediterranean Alps -- Interplay of solar and temperature forcing. Quaternary Science Reviews, 170, 121–135. http://doi.org/10.1016/j.quascirev.2017.06.019

Shipley, B. (2000). Cause and Correlation in Biology, (January 2000). http://doi.org/10.1017/CBO9780511605949

Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. Ecology, 90(2), 363–368. http://doi.org/10.1890/08-1034.1

Steinhilber, F., Abreu, J. A., Beer, J., Brunner, I., Christl, M., Fischer, H., … Wilhelms, F. (2012). 9,400 Years of Cosmic Radiation and Solar Activity From Ice Cores and Tree Rings. Proceedings of the National Academy of Sciences, 109(16), 5967–5971. http://doi.org/10.1073/pnas.1118965109

Steinhilber, F., Beer, J., & Fröhlich, C. (2009). Total solar irradiance during the Holocene. Geophysical Research Letters, 36(19), 1–5. http://doi.org/10.1029/2009GL040142

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14--e14. http://doi.org/10.1093/nar/gkl938

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., … Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology, 21(8), 1816–1820. http://doi.org/10.1111/j.1365-294X.2011.05317.x

Talon, B. (2006). Analyses anthracologiques au col du Petit St-Bernard. Archéoanthracologie et pédoanthracologie. Actes Du Séminaire de Clôture Du Programme INTERREG III A ALCOTRA 2000–2006: Alpis Graia. Archéologie sans Frontière Au Col Du Petit St-Bernard. Aoste, 51–59.

Trimble, S. W., & Mendel, A. C. (1995). The cow as a geomorphic agent--a critical review. Geomorphology, 13, 233–253.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., … Taberlet, P. (2009). New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. Molecular Ecology Resources, 9(1), 51–60. http://doi.org/10.1111/j.1755-0998.2008.02352.x

Vannière, B., Magny, M., Joannin, S., Simonneau, A., Wirth, S. B., Hamann, Y., … Anselmetti, F. S. (2013). Orbital changes, variation in solar activity and increased anthropogenic activities: Controls on the Holocene flood frequency in the Lake Ledro area, Northern Italy. Climate of the Past, 9(3), 1193–1209. http://doi.org/10.5194/cp-9-1193-2013

Vanwalleghem, T., Gómez, J. A., Infante Amate, J., González de Molina, M., Vanderlinden, K., Guzmán, G., … Giráldez, J. V. (2017). Impact of historical land use and soil management change on soil erosion and agricultural sustainability during the Anthropocene. Anthropocene, 17, 13–29. http://doi.org/10.1016/j.ancene.2017.01.002

Vitousek, P. M., Mooney, H. a, Lubchenco, J., & Melillo, J. M. (1997). Human Domination of Earth' s Ecosystems. Science, 277(5325), 494–499. http://doi.org/10.1126/science.277.5325.494

Wilhelm, B., Arnaud, F., Sabatier, P., Crouzet, C., Brisset, E., Chaumillon, E., … Delannoy, J. J. (2012). 1400 years of extreme precipitation patterns over the Mediterranean

French Alps and possible forcing mechanisms. Quaternary Research, 78(1), 1–12. http://doi.org/10.1016/j.yqres.2012.03.003

Wirth, S. B., Glur, L., Gilli, A., & Anselmetti, F. S. (2013). Holocene flood frequency across the Central Alps - solar forcing and evidence for variations in North Atlantic atmospheric circulation. Quaternary Science Reviews, 80, 112–128. http://doi.org/10.1016/j.quascirev.2013.09.002

Zeileis, A. (2003). Testing for structural change in a linear simultaneous equations model under limited information.

Zuazo, V. H. D., & Pleguezuelo, C. R. R. (2008). Soil-erosion and runoff prevention by plant covers . A review. Agron. Sustain. Dev. 28, 28(2008), 65–86. http://doi.org/10.1051/agro:2007062

4.9 Supporting information

4.9.1 Supplementary discussion on plant community change

The statistical analyses performed here provide a rough picture of the potential interactions among climate, human activities, vegetation, and soil erosion. Looking into the long-term evolution of plant communities may draw more details to the picture, in order to have a better understanding of the complex mechanisms involved. The plant DNA dataset indicates dominating pine-forest vegetation through Phase A. This result is congruent with the previously established Alpine treelines of *Pinus cembra*, drawn above at least 2250m a.s.l. until ca. 2700 cal. yr BP (Ali, Carcaillet, Talon, Roiron, & Terral, 2005; Nicolussi, Kaufmann, Patzelt, Plicht Van Der, & Thurner, 2005), and can be viewed as a result of the forest expansion during the Holocene Thermal Maximum from 7000 to 5000 cal. yr BP (Schwörer et al., 2014). Similar to the results of previous SedDNA reconstruction of past vegetation in another subalpine lake catchment (Lake Anterne, Giguet-Covex et al., 2014; Pansu et al., 2015), the most striking transition of the Lake Verney vegetation was the forest clearance and the subsequent expansion of grazed meadows (from Phase A to Phase B, Figure 4-2). In the case of Lake Anterne, however, the opening of landscape in the Lake Anterne catchment was dated as early as 5000-4600 cal. yr BP, in contrast to 2600 cal. yr BP in the case of Lake Verney (Figure 4-2). In the latter catchment area, a drop of *Pinus* DNA did appear in the interval of 4800-5100 cal. yr BP (Figure 4-2, Figure 4-S3), which may be linked to the *Pinus cembra* charcoal dated to the same period, found in the soil at a

nearby site at a slightly lower altitude (1940 m a.s.l) by Talon, 2006. Therefore, a regional wave of high-altitude deforestation during this period by the Neolithic habitants can be hypothesized. In the Lake Verney catchment area, such transition was most probably anthropogenic, because *Sporormiella* spore flux increased during the last stage of Phase A (Figure 4-2), and this index was a strong predictor of PC1 of the vegetation DNA data matrix, which characterized the forest-meadow status.

Vegetation Phase B (2619~1027 cal. yr BP) covers the Iron Age, the Roman period and the early medieval time. This phase was of varying climate conditions that could have been driving both the evolution of vegetation and soil erosion. First, the pine forest was almost completely replaced by meadows during the Iron Age (2700-2000, Figure 4-2). However, the erosion rates in this period were consistently below those predicted by the SEM (Figure 4-5). To this, we propose a climatic explanation: a dryer climate or less frequent extreme precipitation dominating this period led to reduced flood frequencies and/or intensities (Arnaud et al., 2016), which subsequently may translate to less-than-expected soil erosion. Evidences to this explanation can be found in the presence of plant taxa that are indicators of dry soil or atmosphere conditions: e.g. Hedysareae and *Onobrychis montana* (sainfoin, see Figure 4-S3). Besides, XRF analysis of the sediment core found also minimal manganese (Mn) oxide content during this period (Bajard et al., 2017). The precipitation of Mn in lake sediment is mainly associated to water oxygenation, which comes with water precipitation to the lake. Therefore, the reduced Mn peaks during the Iron Age could also indicate less precipitation. Unfortunately, there are few paleoclimatological records on the wet/dry conditions in the Alps for this period to validate our explanation.

During the period of the Roman Empire, the vegetation was characterized by the rise of tall forb community, at the expense of meadow taxa (Figure 4-S3). We propose two mutually complementary causes to this change: 1) the eutrophication of soil due to livestock farming, testified by the peaks of *Rumex* (sorrels, probably *Rumex acetosa*, the field sorrel) DNA (Figure 4-S3), and 2) the wetter and warmer climate during the Roman time (see the mostly

positive TSI during this period in Figure 4-2, see also Büntgen et al., 2011; Holzhauser, Magny, & Zumbuhl, 2005; Simonneau et al., 2014). Besides, the correlation between vegetation PC2 and TSI supports the latter, because PC2 mainly represent the development of tall forb communities (Figure 4-3). In fact, both conditions are known to be favorable for the development of tall forb communities (referred to as 'mégaphorbiaie' in french, Ozenda, 1985).

The tall forb community declined at the beginning of the subsequent Early Middle Ages (ca. 1500-1300 cal. yr BP), which coincided with a minimum of TSI, a progression of the Great Aletsch glacier, and also corresponded to a peak of erosion rate (Figure 4-2). This is in phase with a widespread cold episode i.e. the Dark Age Cold Period (Helama, Jones, & Briffa, 2017). Summer temperature reconstruction based on the Chironomid record of Lake Anterne also revealed a colder climate in the Northwestern Alps during this period (Millet et al., 2009). Therefore, a climatic origin of the change in both the plant communities and the erosion rate may be proposed. The later period of the early Middle Ages was relatively warmer (Millet et al., 2009), with a generally positive TSI, corresponding to a recovery of the tall forb community and a drop of erosion rate in ca. 1300-1200 cal. yr BP (Figure 4-2). However, the erosion rate rose again in 1100 cal. yr BP, probably associated to an intensified livestock farming practice, as we found a peak of *Sporormiella* spore flux and also DNA of sheep in the sample dated to1065 cal. yr BP (Figure 4-2). Interestingly, there was likely a recovery of pine forest during the whole early Middle Ages, accompanied by the regeneration of acid soils typical to pine forests (Bajard, Poulenard, Sabatier, Etienne, et al., 2017), despite the varying climate and human disturbances. This observation may suggest a different herding practice of the people in the early Middle Ages compared to those in the Roman time.

A second expansion wave of grazed meadows replacing the tall forbs and pines (mainly characterized by the vegetation PC1) marked the second major transition of vegetation, arrived at ca. 1000 cal. yr BP. The vegetation remained dominated by meadow herbs

through Phase C until the Modern time. Corresponding to this was the intensified livestock farming indicated by doubled or even tripled *Sporormiella* spore flux, and the simultaneous presence of cattle and sheep (Fig_STR). On the other hand, the first few centuries of Phase C fell into the Medieval Climate Anomaly, a period of favorable warm climate, with a 2°C warming in the Northwestern Alps (Millet et al., 2009) and a retreat of Alpine glacier (Holzhauser et al., 2005), which should have favored the growth of both tall forbs and pine trees. It is therefore most likely that such changes in vegetation are results of the intensified livestock farming, as the SEM analysis suggests (Figure 4-5). Nevertheless, the erosion rate continued to increase and to be responsive to climate variabilities through Phase C, especially during the LIA (Badino et al., 2018). Furthermore, the rapid drop of erosion rate after its second breakpoint at 149 cal. yr BP matched the warming trend after the LIA. These results suggest a climate-driven pattern of soil erosion under heavy livestock grazing.

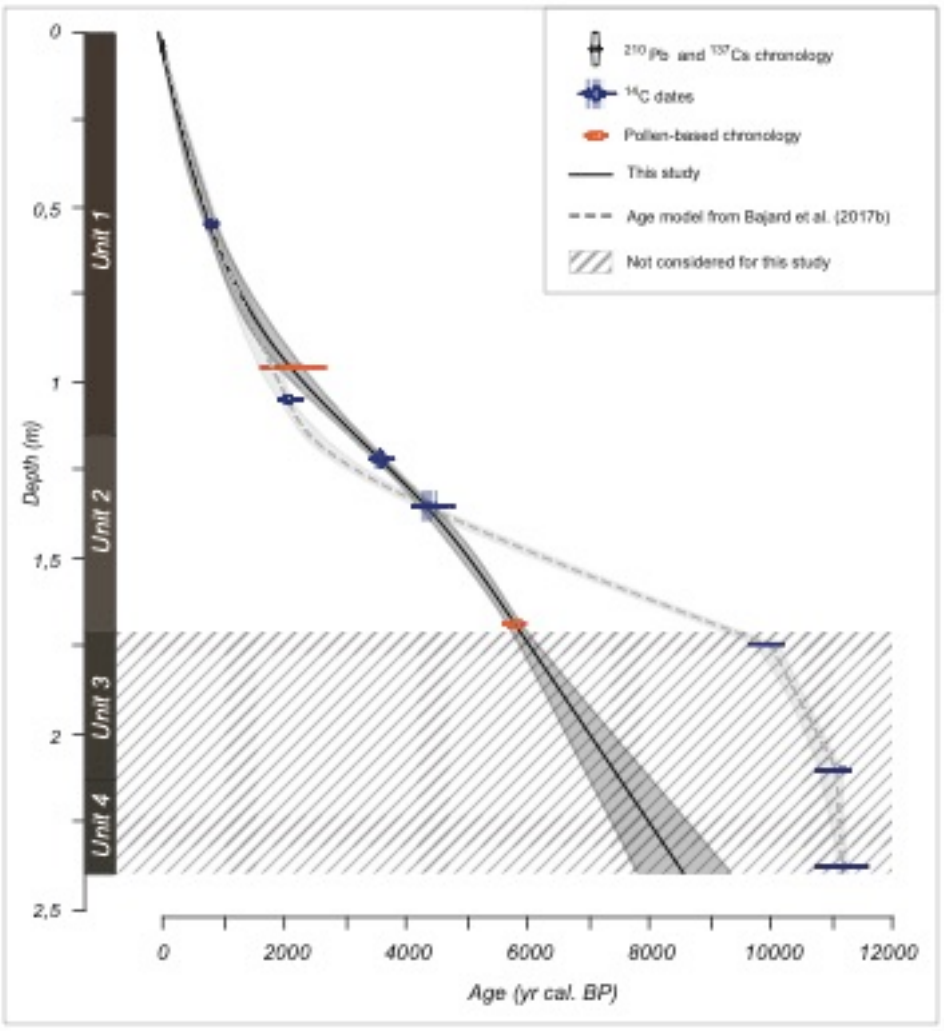


Figure 4-S1: Age-depth model of the Lake Verney sediment sequence modified from Bajard et al. (2017b), based on 14 C, pollen and ^{210}Pb/^{137}Cs ages.

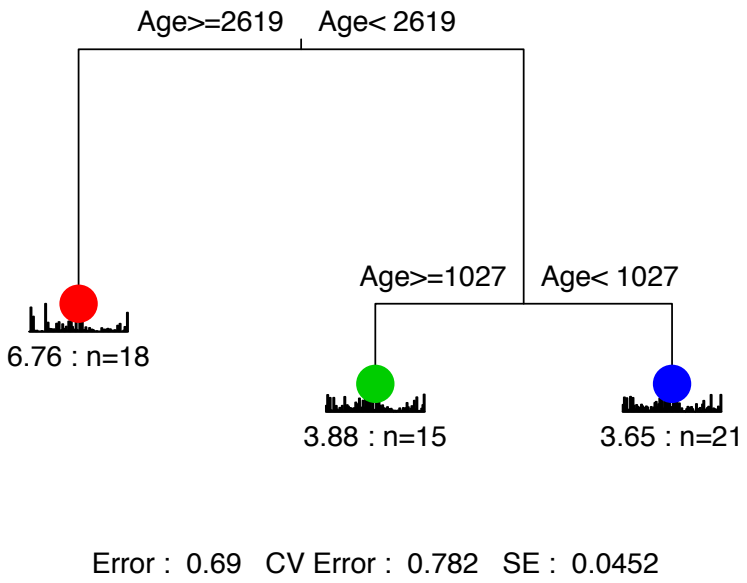
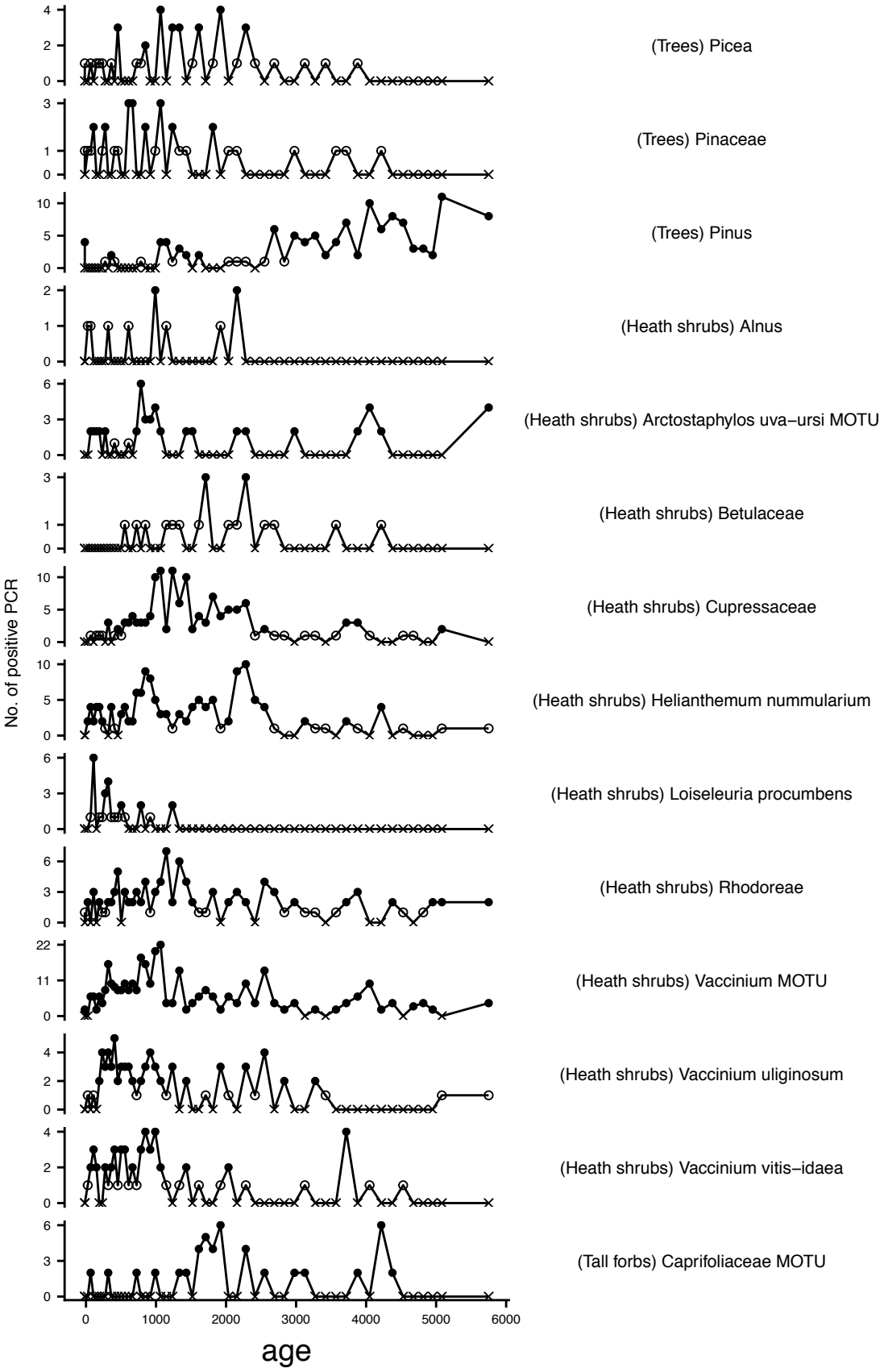
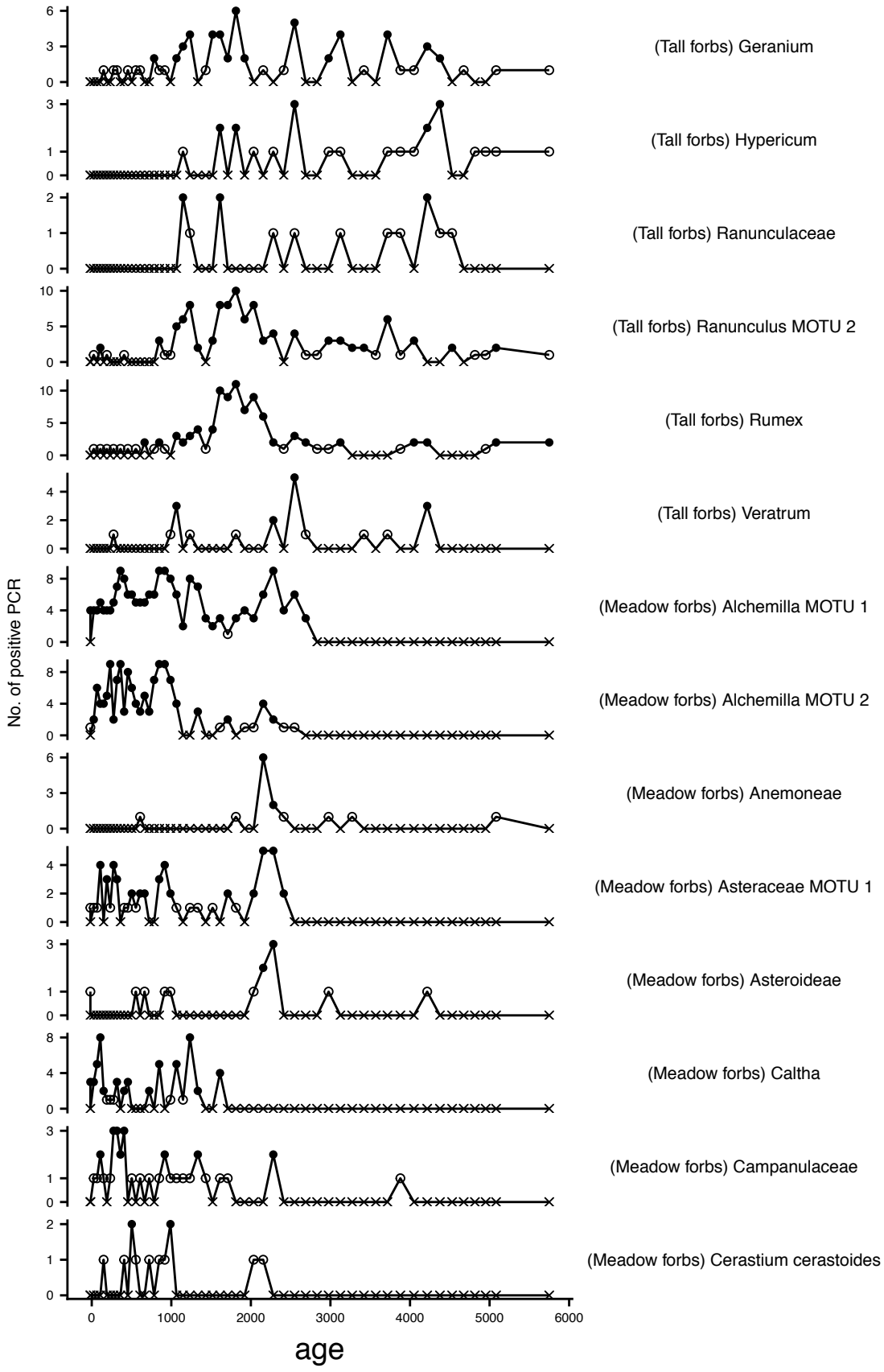
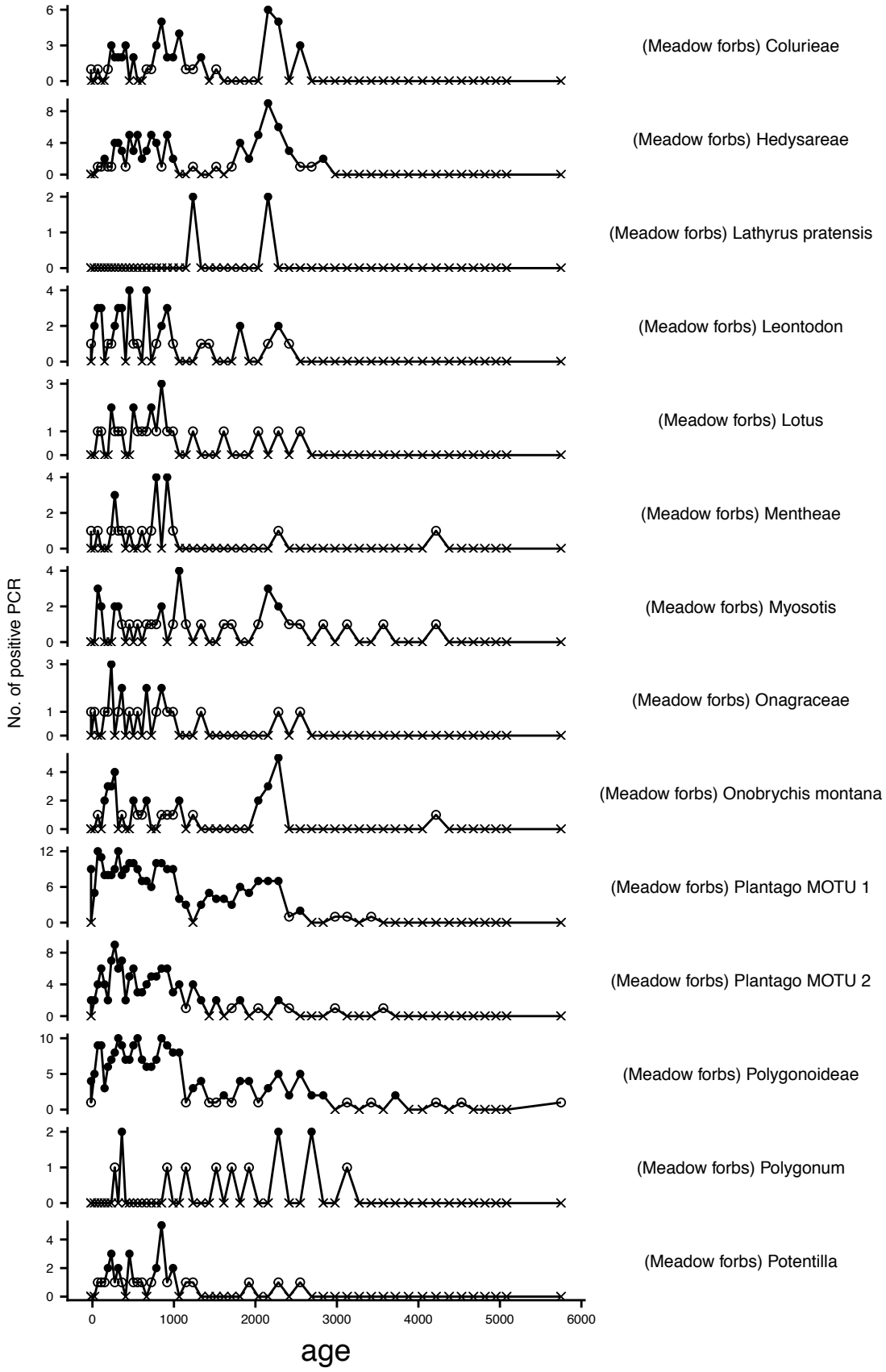
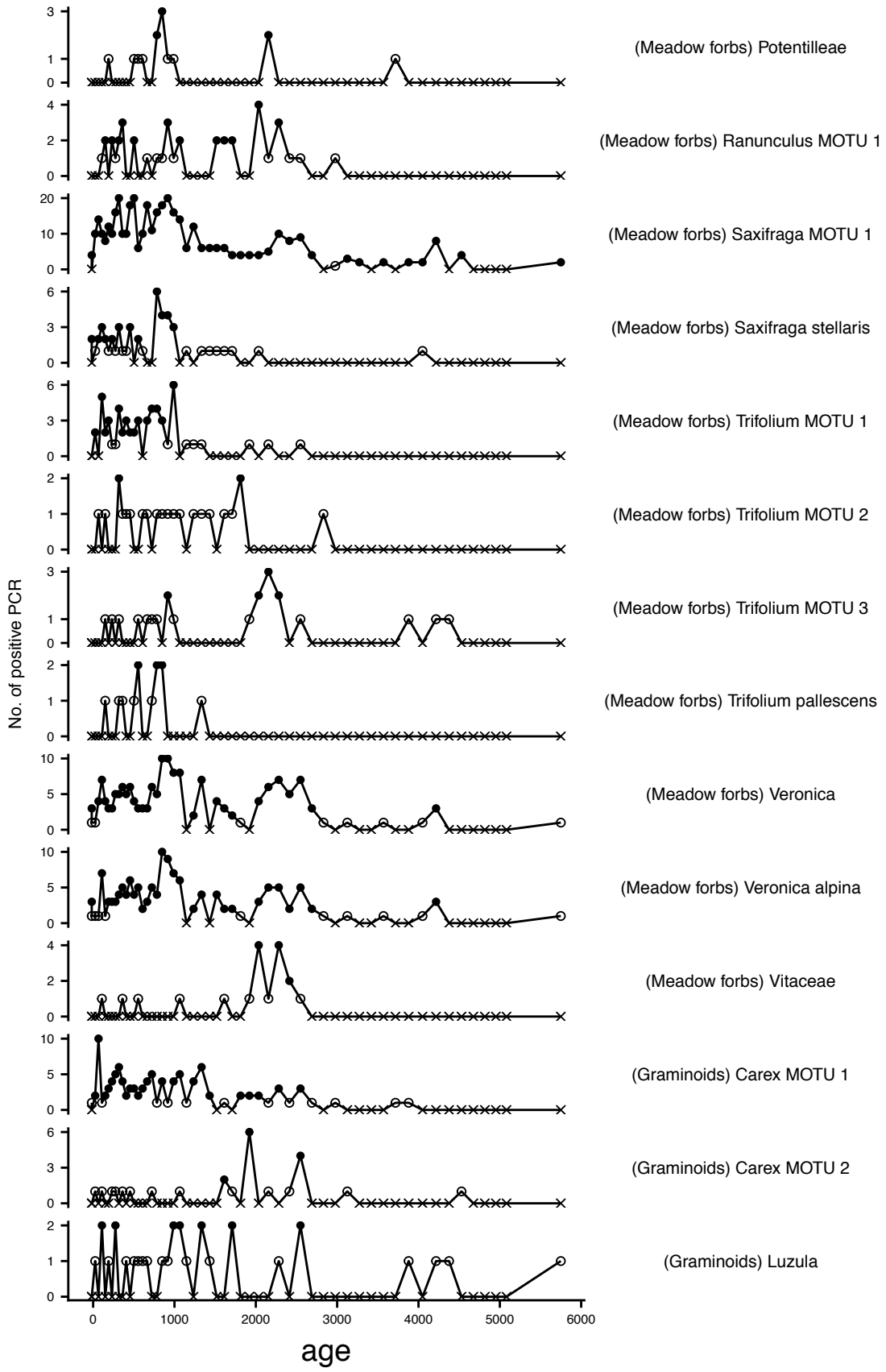


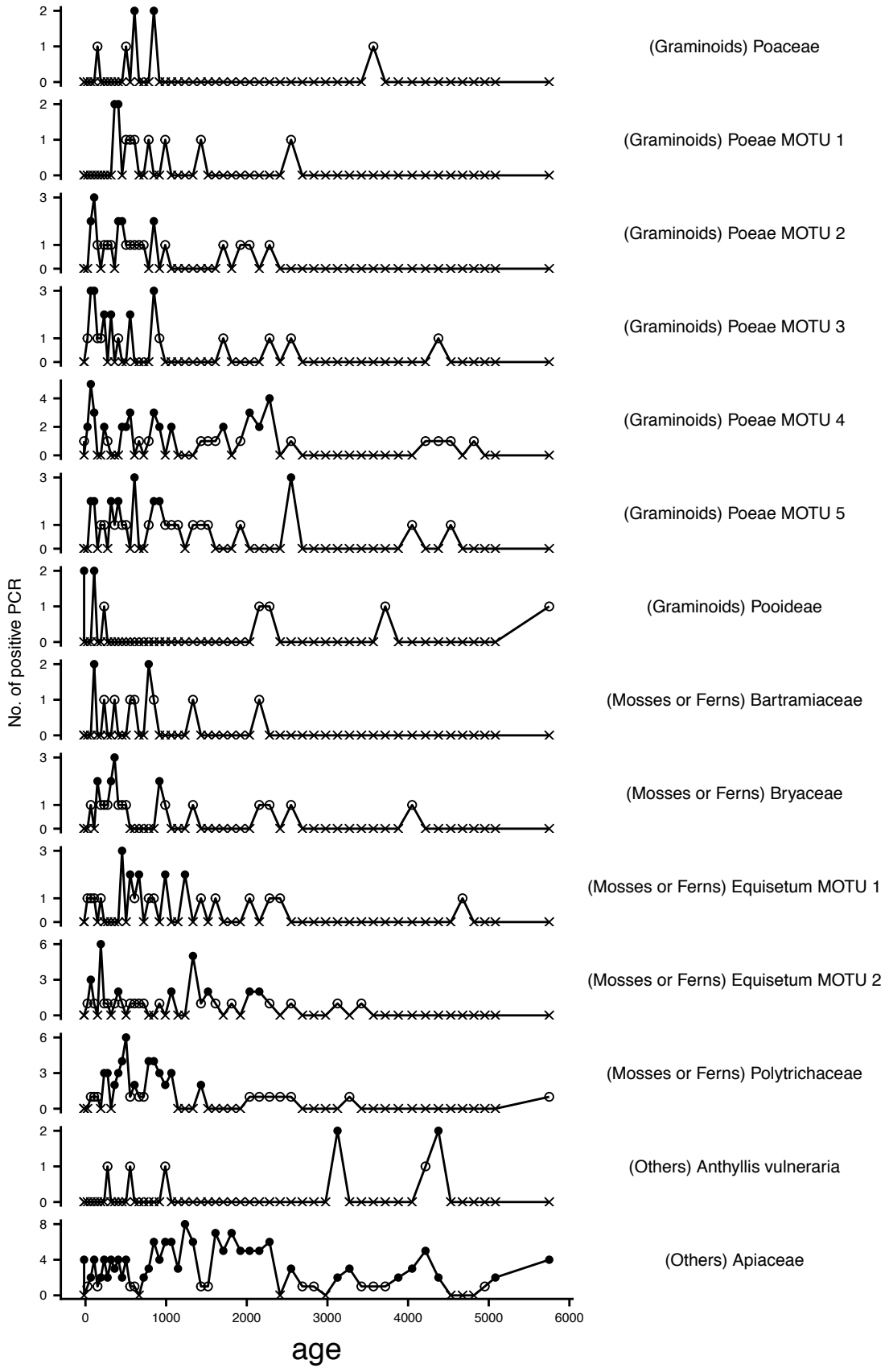
Figure 4-S2: Multivariate regression tree (MRT) on erosion rate in relation with age (cal. yr BP).


No. of positive PCR
(Trees) Picea
(Trees) Pinaceae
(Trees) Pinus
(Heath shrubs) Alnus
(Heath shrubs) Arctostaphylos uva−ursi MOTU
(Heath shrubs) Betulaceae
(Heath shrubs) Cupressaceae
(Heath shrubs) Helianthemum nummularium
(Heath shrubs) Loiseleuria procumbens
(Heath shrubs) Rhodoreae
(Heath shrubs) Vaccinium MOTU
(Heath shrubs) Vaccinium uliginosum
(Heath shrubs) Vaccinium vitis−idaea
(Tall forbs) Caprifoliaceae MOTU
0
1000
2000
3000
4000
5000
6000
age


No. of positive PCR
(Tall forbs) Geranium
(Tall forbs) Hypericum
(Tall forbs) Ranunculaceae
(Tall forbs) Ranunculus MOTU 2
(Tall forbs) Rumex
(Tall forbs) Veratrum
(Meadow forbs) Alchemilla MOTU 1
(Meadow forbs) Alchemilla MOTU 2
(Meadow forbs) Anemoneae
(Meadow forbs) Asteraceae MOTU 1
(Meadow forbs) Asteroideae
(Meadow forbs) Caltha
(Meadow forbs) Campanulaceae
(Meadow forbs) Cerastium cerastoides
0
1000
2000
3000
4000
5000
6000
age


No. of positive PCR
(Meadow forbs) Colurieae
(Meadow forbs) Hedysareae
(Meadow forbs) Lathyrus pratensis
(Meadow forbs) Leontodon
(Meadow forbs) Lotus
(Meadow forbs) Mentheae
(Meadow forbs) Myosotis
(Meadow forbs) Onagraceae
(Meadow forbs) Onobrychis montana
(Meadow forbs) Plantago MOTU 1
(Meadow forbs) Plantago MOTU 2
(Meadow forbs) Polygonoideae
(Meadow forbs) Polygonum
(Meadow forbs) Potentilla
0
1000
2000
3000
4000
5000
6000
age


No. of positive PCR
(Meadow forbs) Potentilleae
(Meadow forbs) Ranunculus MOTU 1
(Meadow forbs) Saxifraga MOTU 1
(Meadow forbs) Saxifraga stellaris
(Meadow forbs) Trifolium MOTU 1
(Meadow forbs) Trifolium MOTU 2
(Meadow forbs) Trifolium MOTU 3
(Meadow forbs) Trifolium pallescens
(Meadow forbs) Veronica
(Meadow forbs) Veronica alpina
(Meadow forbs) Vitaceae
(Graminoids) Carex MOTU 1
(Graminoids) Carex MOTU 2
(Graminoids) Luzula
0
1000
2000
3000
4000
5000
6000
age


No. of positive PCR
(Graminoids) Poaceae
(Graminoids) Poeae MOTU 1
(Graminoids) Poeae MOTU 2
(Graminoids) Poeae MOTU 3
(Graminoids) Poeae MOTU 4
(Graminoids) Poeae MOTU 5
(Graminoids) Pooideae
(Mosses or Ferns) Bartramiaceae
(Mosses or Ferns) Bryaceae
(Mosses or Ferns) Equisetum MOTU 1
(Mosses or Ferns) Equisetum MOTU 2
(Mosses or Ferns) Polytrichaceae
(Others) Anthyllis vulneraria
(Others) Apiaceae
0
1000
2000
3000
4000
5000
6000
age

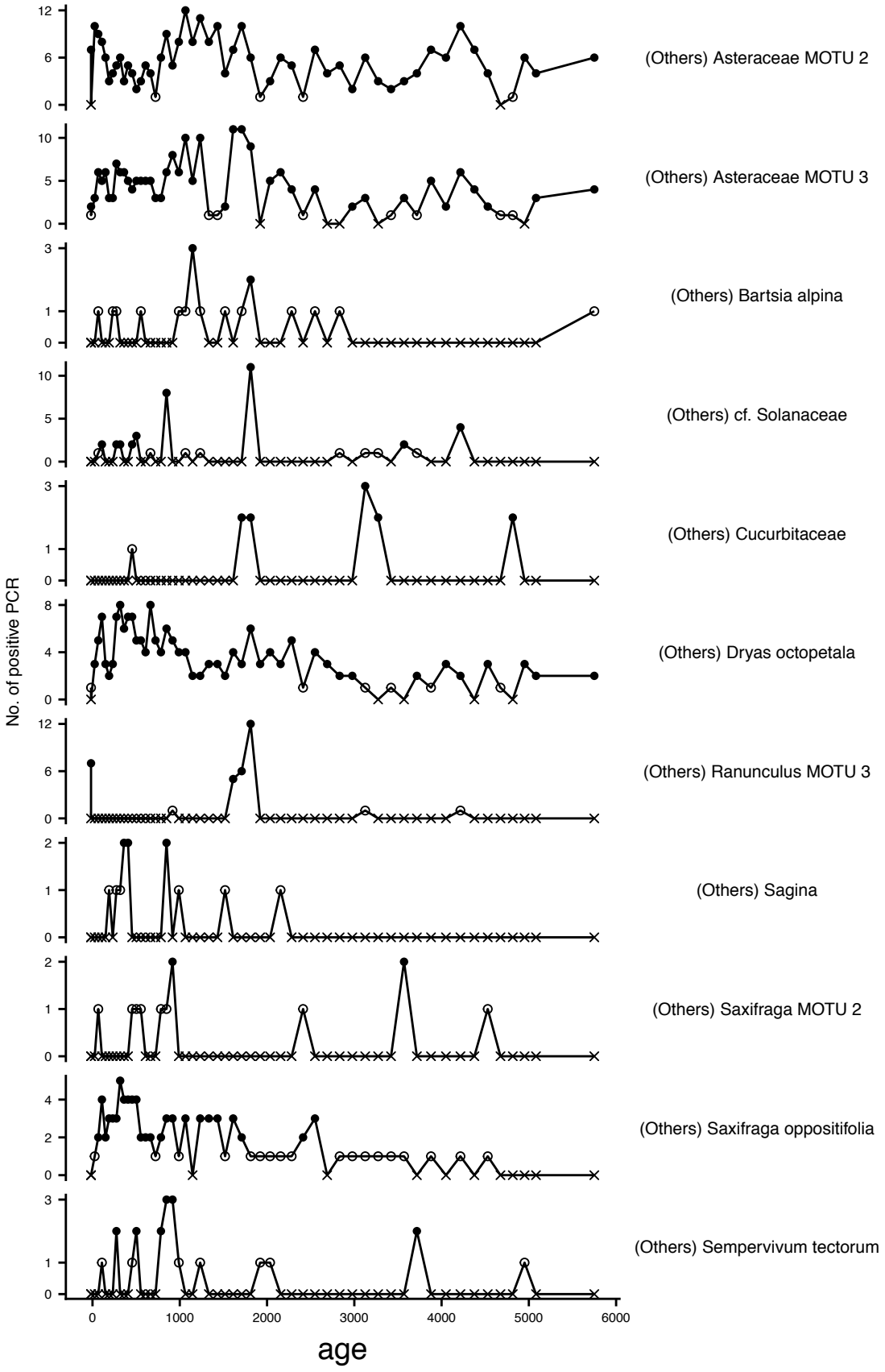

No. of positive PCR
(Others) Asteraceae MOTU 2
(Others) Asteraceae MOTU 3
(Others) Bartsia alpina
(Others) cf. Solanaceae
(Others) Cucurbitaceae
(Others) Dryas octopetala
(Others) Ranunculus MOTU 3
(Others) Sagina
(Others) Saxifraga MOTU 2
(Others) Saxifraga oppositifolia
(Others) Sempervivum tectorum
0
1000
2000
3000
4000
5000
6000
age

Figure 4-S3: chronosequences of numbers of plant DNA positive PCR replicates.

Chapter 5. Conclusions and perspective

The studies presented in the previous chapters demonstrate that sedimentary DNA metabarcoding is a powerful and efficient approach to investigate past environmental changes, and provide key insights on both biodiversity and ecosystem functioning. The combination of sedimentary DNA's local origin and its excellent taxonomic resolution offered unprecedented opportunities to investigate the variability of past community composition (**Chapter 3**) and the complex interactions among climate, human, vegetation and soil that governed the functioning of high-altitude ecosystem functioning (**Chapter 4**).

On the other hand, some of its limitations can be perceived through these studies. First, the degradation of DNA sequences along time prevents unbiased comparisons of community composition among samples of very different ages. Such degradation is most evident by comparing the compositional variability of plant DNA of different periods in **Chapter 3** (Figure 3-4), which showed a generally increasing trend through Periods A, C and D (the higher variability in Period B being interpreted as an 'anomaly' caused by livestock farming), probably resulted from the decreasing detectability of taxa along time related to DNA degradation. It would be possible to alleviate this issue with some alternative DNA-metabarcoding approaches and proper numerical data treatments. For instance, barcode DNA hybridization capture (Shokralla et al., 2014) can largely enrich the target DNA sequences so that the detectability of most taxa could be assured through a long period of time, whereas a dissimilarity measure based on relative abundance [e.g. the Renkonen dissimilarity, (Jost et al., 2011)] could avoid the bias in quantifying DNA abundance introduced by degradation. Besides, it is possible to extend existent SODMs (**Chapter 2**) by taking into account for decreasing detectability along time, providing a more adequate basis of occupancy inference from sedimentary DNA data.

Second, the taxonomic resolution of universal primers is not constant across different groups of organisms. In the case of the *Sper*01 (or *gh*) primers, taxa in some of the most important plant families in the study area such as the Asteraceae and the Poaceae cannot be

effectively identified to the genus level, whereas genus- or even species-level resolution can be reached in some other families such as Ericaceae (see **Chapters 3** and **4**). Unevenness of taxonomic resolution may lead to biased assessment of community composition and miss some crucial information about the system in question. Fortunately, specific primers exist for many important groups of plants, e.g. Asteraceae (*Aste*01), Poaceae (*Poac*01) and Apiaceae (*Apia*01), with relatively short sequence lengths that are required for analyzing ancient DNA (Taberlet et al., 2018). Therefore, multiplexing an universal (e.g. *Sper*01) primer pair with some specific ones depending of the purpose (De Barba et al., 2014) could be a viable solution to this issue. Furthermore, upcoming regional comprehensive reference genome databases (e.g. the Phyloalps project http://phyloalps.org/) can also help solve the problems related to taxonomic resolution.

In the present thesis, we demonstrated how the combination of sedimentary DNA metabarcoding data and existent numerical analysis methods can provide invaluable insights to past environmental changes and human-environment interactions, which could in turn shed light on today's biodiversity and ecosystem functioning issues that are crucial to human well-being. With future advancements in DNA analyzing technology, sedimentary DNA metabarcoding will be able to tackle even more challenging problems. For example, with the cost of sequencing dropping rapidly in the past few years, it is already possible to scale up the breadth of sampling (Porter & Hajibabaei, 2018), by analyzing sedimentary DNA (possibly with multiple markers) extracted from sediment cores taken from dozens of sites in a certain region. This will allow a regional assessment of past environmental change with high spatial resolution, as well as providing unprecedented opportunities to test some fundamental theories in ecology with long-term records (Jackson & Blois, 2015). However, such regional integration would require a better understanding of the origin, transportation, and preservation of sedimentary DNA, in order to choose, and if needed, to develop adequate quantitative analysis methods for normalizing and homogenizing data originated from different sites.

Finally, as DNA analysis technology will continue advancing rapidly in the future, there will be constant needs to adjust existing numerical methods, and to develop new ones, in order to effectively deal with the increasing complicated data acquired from those new technologies. The present thesis is an attempt of developing and optimizing analytical approaches for the analysis of sedimentary DNA. However, a deeper understanding on the origin (and taphonomy) of sedimentary DNA, as well as on the technical limits of the DNA analysis method (e.g. potential biases, relation to biomass) are still crucial to adjust numerical methods for correctly and efficiently extracting useful information from data. To meet this need, cooperation among sedimentologists, molecular biologists, ecologists, and statisticians is indispensable.

Reference

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. Molecular Ecology Resources, 14(2), 306–323. http://doi.org/10.1111/1755-0998.12188

Jackson, S. T., & Blois, J. L. (2015). Community ecology in a changing environment: Perspectives from the Quaternary. Proceedings of the National Academy of Sciences, 112(16), 4915–4921. http://doi.org/10.1073/pnas.1403664111

Jost, L., Chao, A., & Chazdon, R. L. (2011). Compositional similarity and beta diversity. Biological Diversity: Frontiers in Measurement and Assessment, 66–84.

Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology, 27(2), 313–338. http://doi.org/10.1111/mec.14478

Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding: Using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources, 14(5), 892–901. http://doi.org/10.1111/1755-0998.12236

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Abstract

Biodiversity and ecosystem functioning are crucial ecological properties that impact human well-being. Studies on how both properties are affected by human activities and by climate change provide indispensable knowledge to guide natural resource management. Long-term retro-observational data allow to reconstruct past environmental history and offer excellent opportunities to gain such knowledge. Sedimentary DNA is an emerging tool to reconstruct detailed past biodiversity in catchment level, thanks to its excellent taxonomic resolution and highly localized origins. However, previous studies based on sedimentary DNA rarely utilized the existing rich arsenal of numerical ecological analysis methods, which are developed for various types of ecological data. In the present thesis we reviewed the potential applications of such methods on sedimentary-DNA-based studies. With several example studies, we showed how these methods can maximize the knowledge gained from the analysis of multiproxy datasets that included sedimentary-DNA-, sedimentological- and climate records. Despite some limitations, numerical analysis based on sedimentary DNA combined with traditional proxy records is a powerful tool to unravel complex ecosystemic interactions. Future methodological advancements in both DNA analysis and numerical methods are promising to provide invaluable understanding over the drivers of changes in biodiversity and in ecosystem functioning across large spatial and temporal scales.

Résumé en français

La biodiversité et le fonctionnement des écosystèmes sont des propriétés écologiques essentielles qui ont une incidence sur le bien-être humain. Des études sur la manière dont les deux biens sont affectés par les activités humaines et par le changement climatique fournissent les connaissances indispensables pour orienter la gestion des ressources naturelles. Les données de rétroobservation à long terme permettent de reconstituer l'histoire environnementale passée et offrent d'excellentes opportunités d'acquérir de telles connaissances. L'ADN sédimentaire est un outil émergent permettant de reconstituer la biodiversité passée détaillée au niveau du bassin versant, grâce à son excellente résolution taxonomique et à ses origines très localisées. Cependant, les études antérieures basées sur l'ADN sédimentaire utilisaient rarement le riche arsenal de méthodes d'analyse écologique numérique existantes, développées pour différents types de données écologiques. Dans la présente thèse, nous avons examiné les applications potentielles de telles méthodes sur des études basées sur l'ADN sédimentaire. Avec plusieurs exemples d'études, nous avons montré comment ces méthodes peuvent optimiser les connaissances acquises lors de l'analyse d'ensembles de données multiproxy comprenant des enregistrements sédimentaires d'ADN, de sédimentologie et climatiques. Malgré certaines limitations, l'analyse numérique basée sur l'ADN sédimentaire combinée aux enregistrements de proxies traditionnels est un outil puissant pour démêler les interactions complexes écosystémiques. Les futurs progrès méthodologiques dans l'analyse de l'ADN et les méthodes numériques sont prometteurs pour fournir une compréhension inestimable sur les facteurs de changement de la biodiversité et du fonctionnement des écosystèmes à grande échelle spatiale et temporelle.

Annex

Quaternary Science Reviews 170 (2017) 121–135

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

6-kyr record of flood frequency and intensity in the western Mediterranean Alps – Interplay of solar and temperature forcing☆

Sabatier Pierre [a, *], Wilhelm Bruno [b], Ficetola Gentile Francesco [c, d], Moiroux Fanny [a], Poulenard Jérôme [a], Develle Anne-Lise [a], Bichet Adeline [b], Chen Wentao [c, d], Pignol Cécile [a], Reyss Jean-Louis [a], Gielly Ludovic [c, d], Bajard Manon [a], Perrette Yves [a], Malet Emmanuel [a], Taberlet Pierre [c, d], Arnaud Fabien [a]

[a] EDYTEM, Université Savoie Mont Blanc, CNRS, 7337 Le Bourget du Lac, France
[b] LTHE, Université Grenoble Alpes, 38000 Grenoble, France
[c] Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, F-38000 Grenoble, France
[d] Laboratoire d'Ecologie Alpine, CNRS, F-38000 Grenoble, France

ARTICLE INFO

Article history:
Received 15 December 2016
Received in revised form
9 June 2017
Accepted 22 June 2017

Keywords:
Holocene
Paleoclimatology
Western mediterranean alps
Sedimentology
Flood frequency and intensity
Lake sediment
Human activity
Ancient DNA
Earthquakes

ABSTRACT

The high-resolution sedimentological and geochemical analysis of a sediment sequence from Lake Savine (Western Mediterranean Alps, France) led to the identification of 220 event layers for the last 6000 years. 200 were triggered by flood events and 20 by underwater mass movements possibly related to earthquakes that occurred in 5 clusters of increase seismicity. Because human activity could influence the flood chronicle, the presence of pastures was reconstructed through ancient DNA, which suggested that the flood chronicle was mainly driven by hydroclimate variability. Weather reanalysis of historical floods allow to identify that mesoscale precipitation events called "East Return" events were the main triggers of floods recorded in Lake Savine. The first part of this palaeoflood record (6–4 kyr BP) was characterized by increases in flood frequency and intensity in phase with Northern Alpine palaeoflood records. By contrast, the second part of the record (i.e., since 4 kyr BP) was phased with Southern Alpine palaeoflood records. These results suggest a palaeohydrological transition at approximately 4 kyr BP, as has been previously described for the Mediterranean region. This may have resulted in a change of flood-prone hydro-meteorological processes, i.e., in the balance between occurrence and intensity of local convective climatic phenomena and their influence on Mediterranean mesoscale precipitation events in this part of the Alps. At a centennial timescale, increases in flood frequency and intensity corresponded to periods of solar minima, affecting climate through atmospheric changes in the Euro-Atlantic sector.

1. Introduction

Storms and floods are the most frequent and costly extreme weather events occurring in Europe and represented 69% of the economic losses caused by natural disasters between 1980 and 2006 (CEA, 2007). In mountain areas, river flooding is one of the most significant natural hazards, causing widespread damage to infrastructure as well as human and economic losses (Gaume et al., 2009). Therefore, robust knowledge about future trends of storms and floods is one of the main scientific challenges for a worldwide sustainable development. This is particularly true for the European Alps, with their recent demographic and touristic development (Beniston and Stephenson, 2004) and because of the high sensitivity of this region to ongoing climate warming. For the next century, a modification of the hydrological cycle is expected that may lead to changes in precipitation regimes and flood hazards. However, the projections of flood risk are still uncertain, mainly because of the complexity of precipitation pattern variations at a regional scale (IPCC, 2013). A recent study simulated a broad-scale reduction of summer precipitation in the Alps by the year 2100 but an increase in extreme precipitation events over high elevation areas associated with increased convective rainfall (Giorgi et al., 2016). A similar pattern was reported for the last two millennia

☆ The Lake Savine flood deposit data reported in this paper are deposited on NOAA Paleo: https://www.ncdc.noaa.gov/paleo/study/22278.

* Corresponding author.
E-mail address: pierre.sabatier@univ-savoie.fr (S. Pierre).

http://dx.doi.org/10.1016/j.quascirev.2017.06.019
0277-3791/

from the geological flood records of the Northern French Alps in which warmer periods correspond to a lower frequency but higher intensity of extreme flood events (Giguet-Covex et al., 2012; Wilhelm et al., 2013). However, the spatial and temporal distribution of flood events may be rather heterogeneous in the European Alps (Beniston et al., 2007; Giorgi et al., 2016). Additionally, the lack of long-term instrument data at high-elevation sites precludes the identification of significant trends (Beniston et al., 2007; Westra et al., 2013). Therefore, effects of modern climate change on the frequency and the intensity of extreme precipitation events are regionally difficult to anticipate. To overcome this issue, geological data offer opportunities to reconstruct long-term chronicles of past flood events well beyond observational records. Among the geological data, lake sediments are particularly suitable because they provide a well-preserved continuous archive of past flood events (Noren et al., 2002; Bøe et al., 2006; Moreno et al., 2008; Wilhelm et al., 2012a, 2013; Vannière et al., 2013; Wirth et al., 2013; Glur et al., 2013; Swierczynski et al., 2013; Corella et al., 2014; Czymzik et al., 2016).

This study focused on the reconstruction of past flood variability in a French-Italian Alpine border region mainly influenced by the Mediterranean climate. Over instrumental period, heavy precipitation events that triggered catastrophic floods were related to either local convective phenomena (i.e., summer thunderstorms) or mesoscale convective systems called “East Return” events strongly linked to the cyclonic activity across the Mediterranean area (Garavaglia et al., 2010; Gottardi et al., 2010). The main objective of this study was to identify regional flood patterns using historical and lacustrine (Lake Savine) archives. We present a well-dates and high-resolution multi-proxy study of a lake sediment core associating sedimentological and geochemistry data. Our approach focused on six challenges: (i) identify event layers at the millimetre scale, (ii) distinguish event layers induced by floods from those induced by mass movements, (iii) reconstruct the intensity of past floods, (iv) assess the potential impact of human activities on the recorded flood signal through ancient DNA, (v) define the type of precipitation events triggering floods by the study of weather reanalysis and (vi) discuss the long term variability of palaeoflood frequency and intensity with regard to past climatic variations.

2. Study area

2.1. Lake Savine and its catchment area

Lake Savine is located in the western Mediterranean Alps at the French-Italian border in the Haute-Maurienne massif (Fig. 1), approximately 90 km south of Lake Anterne (Giguet-Covex et al., 2011 Figs. 1b) and 105 km north of Lake Allos (Wilhelm et al., 2012a, Fig. 1b). The lake is a rectangular-shaped (550 × 175 m) high-elevation system (2447 m a.s.l., N45°10.500, E6°54.821). Its watershed is formed by a cirque of 3.5 km^2 limited in the southeastern part by the French-Italian border and rises to 3310 m a.s.l. It is mainly (90%) composed of crystalline rocks, Permian gneiss and mica schist. The upper northeastern part of the watershed is composed of Mesozoic rock with Triassic quartzite, dedolomitised breccia (locally named cargneule) and evaporites as well as of Jurrassic to Cretaceous marble and calcareous schist (Fig. 1c).

The main stream of the watershed (Torrent de Savine) drains the southeastern part and has built an alluvial plain (Fig. 1c) that suggests high detrital fluxes. The only evidence of past glacial activity is a moraine in the upper southwestern part of the watershed. Detrital inputs from this stream are limited to summer and autumn because the catchment is covered by snow and the lake is frozen from mid-November to mid-June. The Torrent de Savine flows downstream into the Ambin torrent and then joins the Arc river at the village of Bramans 12 km northwest of Lake Savine. The lake is located above the tree line.

2.2. Hydro-climatic setting and historical flood record

The climate of the study area is influenced by the Mediterranean Sea (Durant et al., 2009). Consequently, the Haute-Maurienne massif corresponds to a transition zone of Alpine precipitation patterns in meteorological reanalyse data between the Southern and the central French Alps (Durant et al., 2009). Heavy precipitation events are related to either local convective phenomena (i.e., summer thunderstorms) or mesoscale convective systems called “East Return” events that mainly occur from late spring to autumn, resulting from Mediterranean humid air masses flowing northward into the Po plain and then westward to the Haute-Maurienne Massif (Gottardi et al., 2010; Wilhelm et al., 2016a). Local historical documents (ONF-RTM database, http://rtm-onf.ifn.fr/) have provided evidence of numerous past flood events. This historical information shows that the village of Bramans has been affected by floods of the Ambin torrent and its two main tributaries (Savine and Etache torrents) at least 15 times during the last 150. Those of June 1957 and October 2000, were extreme precipitation events that resulted in significant damage to houses and infrastructure and cause several deaths (Ratto et al., 2003; Arnaud-Fassetta and Fort, 2004).

3. Material and methods

3.1. Lake coring

A 775-cm-long sediment core was extracted in March 2014 from the frozen surface of Lake Savine using an Uwitec piston coring device directly from the ice-covered surface of the lake. The core was retrieved from the deepest part of the lake (7-m depth, N45°10.500, E6°54.821). The sediment sequence SAV14 (IGSN codes refer to an open international database, www.geosamples.org) was composed of two overlapping coring holes, SAV14-01 and SAV14-02, made of three and two sections, respectively. Sections from the second hole were taken following a 1-m shift in depth to ensure a sufficient overlap to provide a continuous record. A short gravity core (SAV14P2) was also taken to provide a well-preserved water–sediment interface that was subsequently correlated to the composite sequence. The cores were split into two halves at the EDYTEM laboratory. Each half-section was described in detail and pictures were taken at a 20-pixel mm^{-1} resolution. Lithological description of the sequence allowed the identification of different sedimentary facies. A composite sediment sequence named SAV14 was built using distinct marker layers from the overlapping sections of parallel holes (SI1).

3.2. Sedimentary analysis and geochemistry

The grain-size distribution on most of the identified event layers was measured with a 5-mm sampling step using a Malvern Mastersizer 2000G laser particle sizer. Ultrasounic was used to dissociate mineral particles and to avoid flocculation. We then used the median (D50) and the coarsest (D90) fractions to characterize interbedded deposits (Passega, 1964; Mulder et al., 2001; Wilhelm et al., 2013, 2015). We also recorded the thickness and D90 max (i.e., the highest D90 value) of each interbedded deposit.

78 samples of 2-cm wide were sampled on specific interval along the SAV14 mastercore. After taking the inner part for ancient DNA analysis, external parts were dried at 60 °C over 4 days to obtain dry bulk densities. Then, the Loss Of Ignition (LOI) of each sample was determined using the protocol of Heiri et al. (2001). The

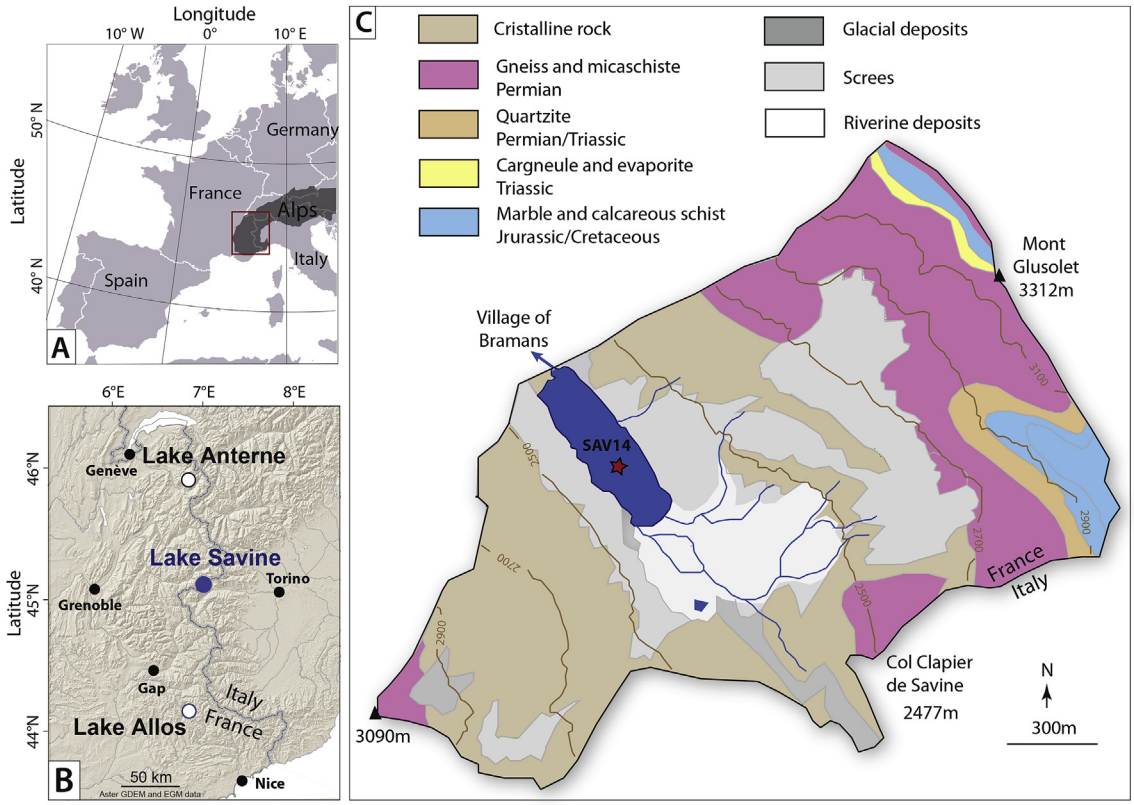


Fig. 1. (A) Location of Lake Savine in the Western Mediterranean Alps, (B) compared with the locations of Lake Anterne (Giguet-Covex et al., 2011) and Lake Allos (Wilhelm et al., 2012a). (C) Geological and geomorphological characteristics of the Lake Savine catchment area.

LOI at 550 °C and 950 °C corresponds to the organic and carbonate components of the sediment, respectively.

X-Ray Fluorescence (XRF) analysis was performed for the entire composite sequence with a 1-mm sampling step using the non-destructive Avaatech core-scanner (EDYTEM). The split core surface was first covered with 4-μm-thick Ultralene foil to avoid contamination of the XRF measurement unit and desiccation of the sediment. The geochemical data were obtained with two tube settings: 10 kV and 1.5 mA for 20 s for Al, Si, S, K, Ca, Ti, Mn, and Fe and 30 kV and 0.8 mA for 40 s for Cu, Zn, Br, Sr, Rb, Zr, and Pb (Richter et al., 2006). Each individual power spectrum was converted by a deconvolution process into relative components (intensities) expressed in counts per second.

3.3. Ancient DNA (aDNA)

The 78 slices of 2-cm thickness were sampled following the strict laboratory precautions described in Giguet-Covex et al. (2014). For each sediment slice, we mixed 15 g of sediment with 15 ml of saturated phosphate buffer (Na2HPO4; 0.12 M; pH ≈ 8) for 15 min. The mixture was then centrifuged (10 min at 10,000 × g), and 12 ml of the resulting supernatant were transferred to Amicon® Ultra-15 10 K Centrifugal Filter Devices (Millipore) and centrifuged (20 min at 4000 × g) for concentration of aDNA. Of the resulting concentrate, 400 μl was kept as starting material for DNA extraction, using the NucleoSpin® Soil kit (Macherey-Nagel) (Taberlet et al., 2012). Four extraction controls were performed. Mammal DNA was amplified with the primer pair Mam-P007 following the protocol described in Giguet-Covex et al. (2014). In addition to extraction controls, we performed three PCR control runs containing PCR mix but no DNA template and four PCR positive controls, each containing 0.18 ng of DNA extracted from a marsupial (*Dideplhis marsupialis*) absent in Europe. All samples and controls were amplified in 12 replicated PCRs (Ficetola et al., 2015). Sequencing was performed by 2 × 125 bp paired-end sequencing on the Illumina HiSeq 2500 platform, which returned 21,192,000 reads.

DNA sequences were filtered using the OBITOOLS software (Boyer et al., 2016) as described in Pansu et al. (2015). Sequences were assigned to the relevant taxa by comparing them with a global mammal database generated from EMBL using the *ecotag* programme. We only kept sequences with a match >97% with a mammal genus. In aDNA studies, false detections (i.e., sporadic detections of absent species) are possible. Therefore, we only considered species confirmed by multiple PCR analyses performed on the same sample, and we discarded sequences detected with <5 reads in each PCR replicate and sequences frequently detected in controls (humans and pig). We used site occupancy-detection modelling to estimate the detection probability of species and the frequency of false detections (Ficetola et al., 2015). Occupancy modelling was performed using Bayesian models, following Lahoz-Monfort et al. (2016). Estimates of occupancy, detection probability and the probability of false detection were then used to calculate the probability of species presence in each sample. This approach allowed the successful limiting of false positives in aDNA analyses, and the probability of presence was a measure of the reliability of species presence in a sample (Ficetola et al., 2016; Lahoz-Monfort et al., 2016).

3.4. Chronology

The chronology of the Lake Savine sediment sequence was

based on short-lived radionuclide measurements and 12 ^{14}C measurements on terrestrial macro-remains. The short-lived radionuclides in the upper 12 cm of core SAV14P2 were measured using high-efficiency, very low-background, well-type Ge detectors at the Modane Underground Laboratory (LSM) (Reyss et al., 1995). The measurement intervals followed facies boundaries and resulted in a non-regular sampling of approximately 5 mm. The isotope ^{137}Cs was accidentally introduced into the environment at the end of the 1950s as by-product of atmospheric nuclear weapons tests (peak at AD 1963). The Chernobyl accident in 1986 also dispersed ^{137}Cs into the northern atmosphere (Appleby, 1991). ^{210}Pb excess was calculated as the difference between total ^{210}Pb and ^{226}Ra activities (Golberg, 1963). We then used the Constant Flux/Constant Sedimentation (CFCS) model and the decrease in excess ^{210}Pb to calculate sedimentation rates (Golberg, 1963). The uncertainty of sedimentation rates obtained by this method was derived from the standard error of the linear regression of the CFCS model. The ^{14}C measurements were carried out by an Accelerator Mass Spectrometer (AMS) at the Laboratoire de Mesure ^{14}C (LMC14) ARTEMIS at the CEA (Atomic Energy Commission) institute at Saclay (Table 1). The IntCal13 calibration curve (Reimer et al., 2013) was used for ^{14}C age calibration. Then, we used the software-package "clam" (version 3.0.2 R Development Core Team, 2011) to generate an age/depth model (Blaauw, 2010).

3.5. Weather analysis

Weather analyses were realized using the 20th-century reanalysis (version 2) product of Compo et al. (2011), which contains global atmospheric variables since 1871 on a 2° grid (hereafter 20CR). The reanalysis is based on the Ensemble Kalman Filter data assimilation system of Whitaker and Hamill (2002) and assimilates observed surface pressure and sea-level pressure every 6 h. All reanalysis data used in the current study were related to the mean of the 56 ensemble members. The atmospheric circulation patterns associated with historical flood events recorded in Lake Savine were based on a daily mean of 500 hPa geopotential height anomalies (Z500) and vertically integrated water vapor transport (hereafter IWT). IWT was computed from the daily fields of specific humidity (q), zonal (u), meridional (v) winds and differential of pressure (dp). The magnitude of daily IWT was calculated in an Eulerian framework as follows:

$$IWT = \left[\left(\int_{1000}^{300} qu \frac{dp}{g} \right)^2 + \left(\int_{1000}^{300} qv \frac{dp}{g} \right)^2 \right]^{\frac{1}{2}}$$

where g is the acceleration due to gravity. The vertical integration was limited to the 1000 to 300 hPa pressure interval because specific humidity in the 20CR data is negligible above 300 hPa. Through the manuscript D represents the day of historical precipitation event, D-2, D-1 mean two and one days before D respectively.

4. Results

4.1. Core description and lithology

The upper 7 m of sediment consisted of dark to very dark olive silt composed of 5 ± 1.3% of LOI550 (organic matter, OM) and 7.5 ± 1% of LOI950 (Carbonate fraction) representing the background hemi-pelagic sedimentation (named BS for background sedimentation, Fig. 2). Below 7 m the sediment consisted of homogenous light grey clay to silt with 2 ± 0.1% LOI550 and 6.8 ± 0.1% of LOI950. Because no plant macro-remains requested for (radiocarbon) dating were found deeper than 5.7 m, this study is focused on the upper 5.7 m of the Savine sedimentary record. These fine grained deposits were interrupted by 220 graded layers that are interpreted to represent short-term depositional events according to Sturm and Matter (1978) and Arnaud et al. (2002).

The 220 event layers were divided in two types of deposits based on their grain size and geochemical patterns. Type-1 (T1) deposits were light-grey graded bed with no evidence of an erosive base (Fig. 2). The 200 T1 deposits are characterized by a silty-to-sandy base, a central part with a regular decrease of both D50 and D90 and a very light-grey clayey cap (Fig. 2). The bottom of T1

Table 1
^{14}C dates for Lake Savine. Samples in bold correspond to rejected dates.

Samples	Material dated	Cores	Composite depth (mm)	Depth without instantaneous deposits (mm)	Uncalibrated age (BP)	Uncertainty	Calibrated age ranges at 95% confidence interval (cal. BP)
SacA38347	**Organic macroremain**	**SAV02A**	**355**	**225**	**450**	**30**	**473–535**
SacA38348	Organic macroremain	SAV02A	655	275	405	30	329–516
SacA38349	**Organic macroremain**	**SAV02A**	**885**	**350**	**895**	**30**	**737–909**
SacA38350	Organic macroremain	SAV02A	1265	455	645	30	556–668
SacA38351	Organic macroremain	SAV01A2	1520	535	1100	30	938–1062
SacA38352	Organic macroremain	SAV01A2	1775	560	1315	30	1179–1296
SacA38353	Organic macroremain	SAV01A2	2310	679	1740	30	1564–1714
SacA39966	Organic macroremain	SAV01A2	2765	851	1985	30	1878–1994
SacA38354	Organic macroremain	SAV02B1	3465	1036	2525	40	2472–2745
SacA38355	Organic macroremain	SAV01B2	4425	1397	3560	30	3725–3963
SacA38356	Organic macroremain	SAV01B2	5265	1633	4485	35	4980–5294
SacA38357	Organic macroremain	SAV02C1	5655	1866	5260	35	5932–6178

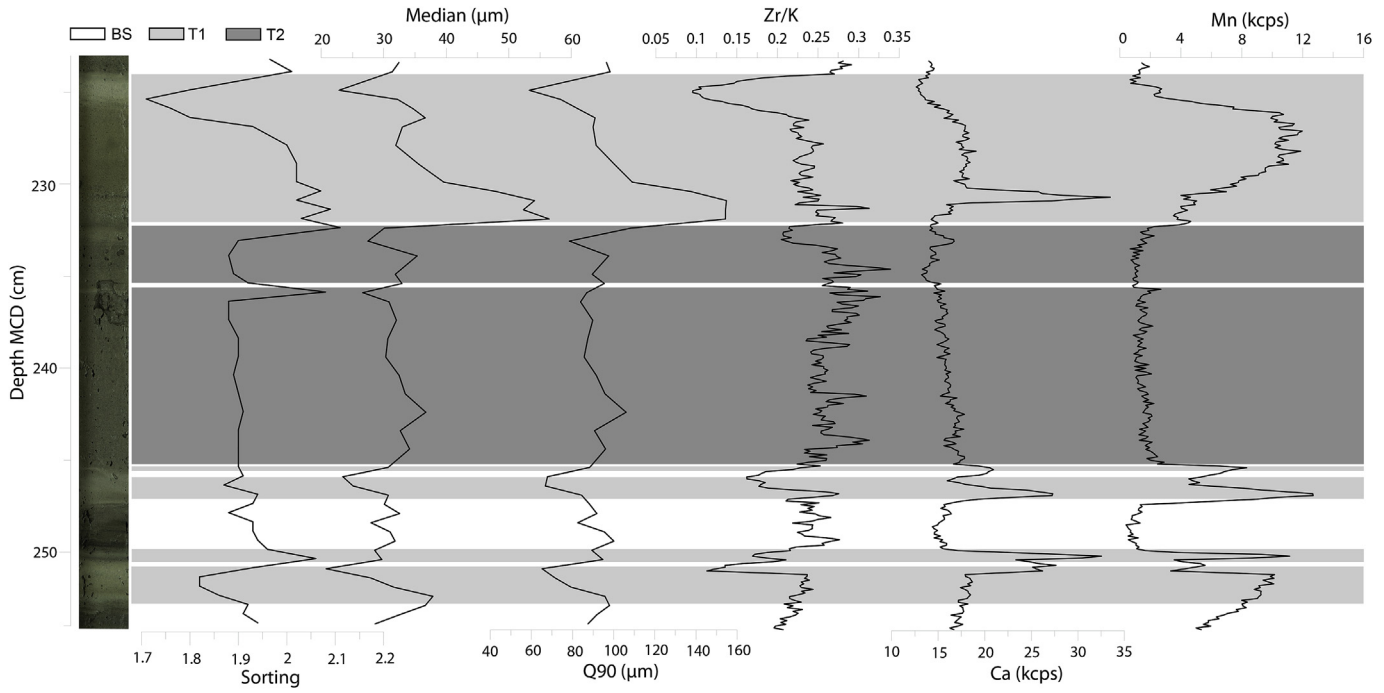


Fig. 2. Photography, grain size and geochemical data examples of the three sedimentary facies of SAV14 (BS: background sedimentation in white, T1 in light grey and T2 dark grey).

deposits are characterized by relatively high sorting values (mostly > 2, Fig. 2). Their geochemical pattern is characterized by relatively high Zr intensities at the base and high K intensities in the clay cap, suggesting Zr enrichment in the coarser sediment fraction and K enrichment in the finest fraction. The Zr/K ratio was thus supposed to be a high-resolution proxy of grain-size (Wilhelm et al., 2013; Arnaud et al., 2016). However, the relation between the Zr/K ratio and grain size parameters is not well constrained, probably in relation to different sample resolutions (1 mm and 5 mm respectively). Thus this geochemical ratio thus cannot be used as a high-resolution grain-size proxy but was used to identify T1 deposits. Ca intensities in these deposits showed high values (Fig. 2) simultaneous with high values of LOI950, which suggests high carbonate contents. This increase in Ca reflects detrital calcareous input from the eastern part of the watershed related to torrential activity (thanks to detailed Scanning Electron Microscopy (SEM) analysis, data not shown). Mn intensities also showed high values in T1 deposits (Fig. 2) that probably reflect the oxygenation state of the water column (Calvert and Pedersen, 1993; Elbaz-Poulichet et al., 2014). The thickness of T1 deposits varied between 2 and 170 mm with a mean of 20 mm (Fig. 3).

Type-2 (T2, Fig. 2) deposits were composed of homogeneous, olive to grey, basal silty layers topped by light grey mm-thick clay caps. Their grain size parameters (median, D90) did not show any significant variations except for the clay cap, where lower values of median and D90 were observed (Fig. 2). The sorting of T2 deposits always presented weak values (approximately 1.9) except for the uppermost samples of each deposit because the clay cap is very thin and thus a mix with the silts of the background sediment. Geochemical data presented a constant and relatively high Zr/K ratio, except for the clay cap, and low values of Ca and Mn intensities (Fig. 2).

Plotting the thickness of all T1 (n = 26) and T2 (n = 16) deposits thicker than 1 cm versus D90 max (Wilhelm et al., 2012a,b) highlights two distinct patterns (Fig. 3). A significant positive linear relationship between bed thickness and D90 max (r = 0.94, $p < 10^{-12}$) was observed for T1 deposits (Fig. 3). This suggests that the grain size of the mobilized sediments is larger when the volume of sediments transported and deposited is higher. For the T2 deposits, a distinct pattern was observed of a high variability of deposit thickness without high variations in D90 max. This suggests that the volumes of sediments transported and deposited to form T2 deposits were poorly related to grain size and thus to underwater current velocities.

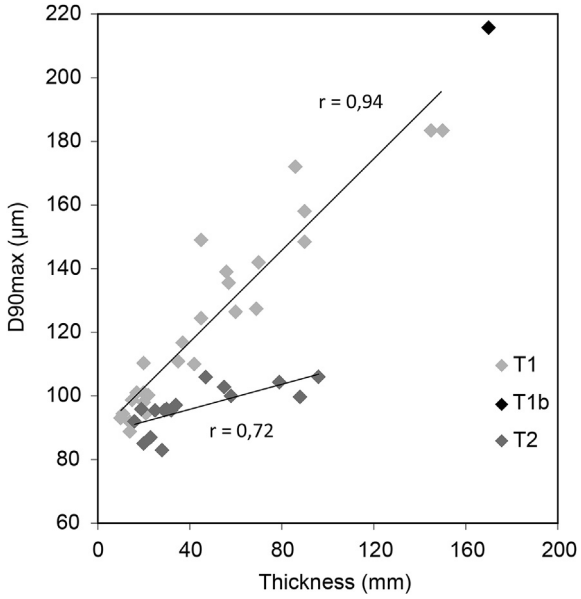


Fig. 3. Diagram of the D90 max-deposit thickness for T1 and T2 deposits. T1b corresponds to the mass movement-induced deposit above the reversed sequence; see SI3.

Finally, a remarkable 22-cm thick reversed sequence was observed between 196 and 218 cm depth of the master core overlapped by a 17-cm thick graded layer (detailed in SI3). This T1

deposit, labelled as T1b in Fig. 3, corresponds to the thickest layer of the entire sequence and was not considered in the comparison of grain size thicknesses. It was localized at the top of the reversal sequence (SI3) and have the same characteristic than other T1 deposits.

4.2. Chronology

4.2.1. Short-lived radionuclides

The excess ^{210}Pb downcore profile showed a regular decrease punctuated by distinct drops (Fig. 4). Following Arnaud et al. (2002), these low values of ^{210}Pbex corresponded to T1 deposits and were excluded from the construction of an event-free sedimentary record because they were considered as instantaneous deposits. ^{210}Pbex activities plotted on a logarithmic scale revealed two linear trends providing two mean sedimentation rates of 1.39 ± 0.2 mm yr^{-1} above 4.5 cm and 0.44 ± 0.04 mm yr^{-1} below. Ages were then calculated using the CFCS model applied to the original sediment sequence to provide a continuous age-depth relationship.

The ^{137}Cs profile presented one peak at 4.5 cm with a high activity of 692 ± 6 Bq.kg^{-1}. This corresponds to the Chernobyl accident in 1986 CE, and the localization of the study site at the border between France and Italy strengthened this interpretation because Italy had larger ^{137}Cs fallout from this accident. The ^{137}Cs activity from maximum nuclear weapon tests in the Northern Hemisphere in 1963 CE presents a regional value between 40 and 300 Bq. kg^{-1} (Wilhelm et al., 2012b; Jenny et al., 2013; Alric et al., 2013; Sabatier et al., 2014; Bajard et al., 2015; Guédron et al., 2016). Such a ^{137}Cs activity was reached at approximately 5.8 cm, even if this ^{137}Cs peak is not clearly identifiable probably in relation to the sample resolution. Both artificial markers are in agreement with the CFCS age model for the last century (Fig. 4).

4.2.2. Radiocarbon dating

A total of 12 terrestrial organic macroremain samples were dated by ^{14}C age (Table 1). Two radiocarbon ages were excluded because they appear to be too old, probably due to macroremain stored in the lake catchment and would result in age inversions (Fig. 5). The 220 deposits interpreted as instantaneous events and the reversed sequence (SI3, noted R in Fig. 5), representing a total of 391 cm were removed. The remaining 200 cm (Fig. 5A) were used to build an event-free sedimentary record (Bøe et al., 2006; Wilhelm et al., 2012a, 2013, Fig. 5A). We then calculated an age-depth relationship by a smooth spline interpolation using the R-code package "Clam" version 2.2 (Blaauw, 2010). This age-depth

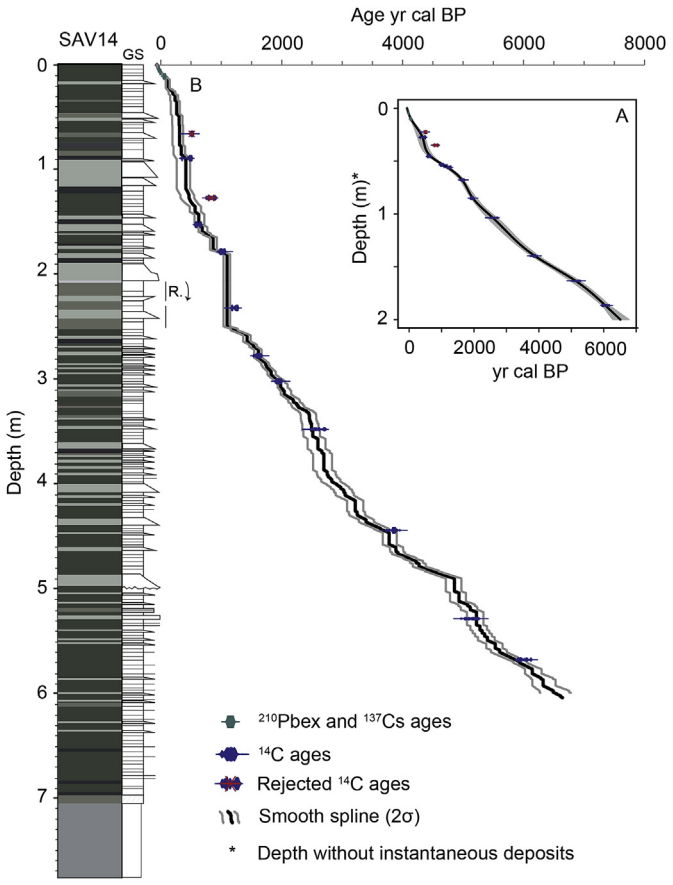


Fig. 5. (A) Age-depth model of the event-free sediment record based on radiocarbon- and short-lived radionuclides-derived ages. (B) Age-depth model for the entire SAV14 sediment sequence. Colour is defined is SI2. R indicates the reversed sequence; see SI3. GS: grain size. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

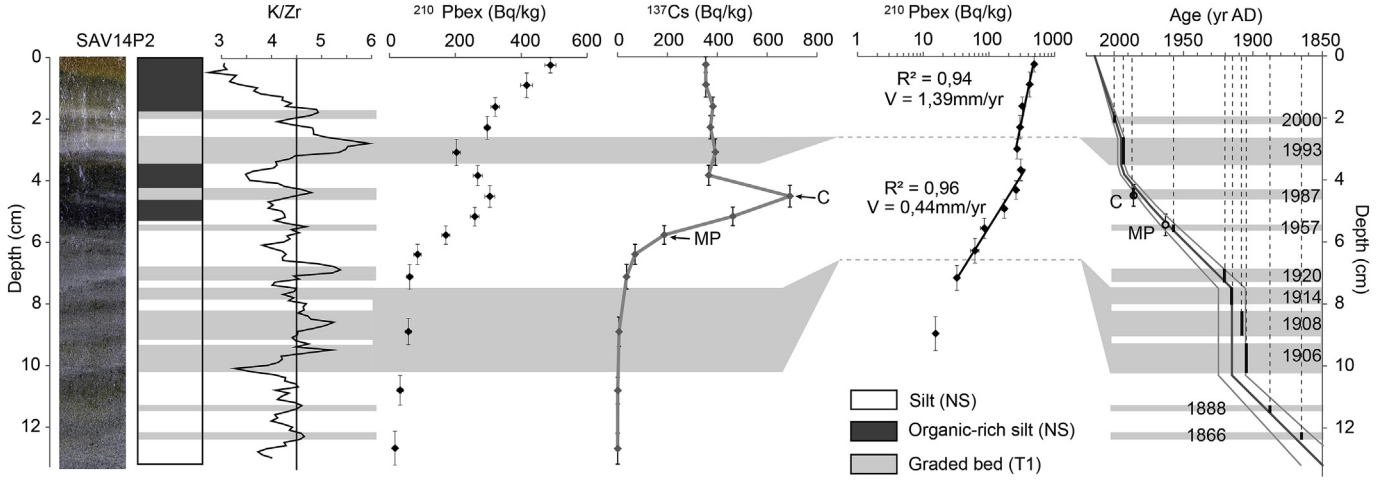


Fig. 4. Chronology (with 1σ uncertainties) of the uppermost part of core SAV14P2 based on short-lived radionuclides (^{210}Pbex and ^{137}Cs) and application of the CFCS model to the event-free ^{210}Pbex profile. Uncertainties of ^{210}Pbex and ^{137}Cs activities are included in dots. Dates on the right-hand side correspond to dates of historical floods, possibly associated with the most recent T1 deposits.

model was used to date all instantaneous deposits. The vertical bars represent the age of each flood thicker than 5 mm with uncertainties (2σ) resulting from the ^{14}C ages (Fig. 5B). The first 565 cm of the sediment sequence from Lake Savine covered the last 5.8 kyr BP. The mean sedimentation rates estimated from the event-free sediment record varied between 0.2 and 0.4 mm yr^{-1} with a short-term increase to 0.6 mm yr^{-1} between 1.95 and 1.65 kyr BP, to 1.4 between 0.35 and 0.55 kyr BP and to 0.44 during the last 40 years (Fig. 6D).

4.3. Ancient DNA

We detected DNA of the taxa of four mammals: *Bos* (cattle), *Canis* (dog), *Felis* (cat) and *Ovis* (sheep). However, only cattle and sheep DNA was detected in more than one sample by multiple PCR replicates. Therefore, we focused the analyses on these two species, whereas we assumed the presence of dog and cat DNA to be sporadic.

But also cattle DNA occurred sporadically in sediments older than 2460 yr cal BP (below 332 cm depth). However, in these cases, sheep DNA was only detected in one of the 12 PCR replicates performed on each sediment slice. Consequently, the reliability of sheep detections was low. Conversely, cattle and sheep DNA was consistently detected between 2050 and 1750 yr cal BP (289 and 307 cm) in multiple PCR replicates per sample, which indicates high reliability of their presence in this short period (Fig. 6E and F) (Giguet-Covex et al., 2014).

5. Discussion

5.1. Event layers

5.1.1. Flood-induced layers

Normally graded layers like T1 deposits are common features of lake sediments, where they are associated with turbidity currents triggered by mass movements or flood events (Sturm and Matter, 1978; Arnaud et al., 2002; Gilbert et al., 2006). In the thickness–D90 max diagram, T1 deposits are grouped in a distinct cluster, which suggests that they resulted from the same trigger mechanisms (Fig. 3). Additionally, the increase of transported and deposited sediment volumes with grain size characterize T1 deposits (Fig. 3) as consistent with sedimentary processes regulated by flowing water currents (e.g., Wilhelm et al., 2013, 2015). Higher discharge rates lead to a larger sediment supply and coarser grains transported to the lake. The 200 T1 deposits are also characterized by increases in Mn, which suggests that changes in redox conditions occurred in the lake system. In oxic waters, Mn occurs as Mn(IV) and diffuses downwards where it can precipitate as MnCO3 or Mn–Ca carbonates (Calvert and Pedersen, 1993; Pedersen and Price, 1982) and upwards where it can precipitate as Mn oxides at the oxic/anoxic transition (Elbaz-Poulichet et al., 2014). This process leads to Mn enrichment at the oxygen-rich water-sediment interface (Calvert and Pedersen, 1993). When bottom waters are anoxic, Mn(II) diffuses to the water column and Mn becomes depleted in the sediments. The presence of Mn in the graded layers associated with Ca (Fig. 2) suggests that hyperpycnal turbidity currents favoured bottom water re-oxygenation (Wilhelm et al., 2016a) and produced rapid precipitation of Mn oxides and/or Mn–Ca carbonates in the sediment (Calvert and Pedersen, 1993).

To verify this hypothesis, we compared our record with the historical flood calendar from the Ambin torrent and its two main tributaries to the event layers archived in the lake sediment for the last 150 years. Ten historical flood events occurred in 2000 CE, 1993, 1987, 1957, 1920, 1914, 1908, 1906, 1888 and 1866 and correspond well to the age of T1 deposits identified by the Zr/K ratio of the uppermost 13 cm (Fig. 4). All these arguments support that T1 deposits were induced by floods.

To reconstruct the intensity of past floods, two proxies could be used; grain size and thickness of the flood deposits. The grain-size data are interpreted to represent the river energy and, thereby, the discharge (e.g. Campbell, 1998; Parris et al., 2010; Lapointe et al., 2012; Wilhelm et al., 2015). The grain-size distribution reflects hydrologic conditions, in particular transport capacity during a flood event, thus the maximum grain size (Q90 max) is a proxy of stream velocity during flood events (Molinaroli et al., 2009; Parris et al., 2010) and it could be applied to assess the intensity of past floods (Gilli et al., 2013). The thickness is interpreted to represent the total volume of solid transported material and deposited during the flood event (Schiefer et al., 2011; Wilhelm et al., 2012a, 2015; Jenny et al., 2014). The significant relationship between D90 max and thickness of T1 flood deposits (Fig. 3) suggests that both parameters could be used, but thickness is more easy measured on each flood layer this is why we choose thickness as a proxy of flood intensity.

5.1.2. Mass movement-induced layers

The thickness-D90 max diagram reveals that T2 deposits resulted from processes distinctly differs from those of T1 deposits. This diagram also suggests that the volume of transported and deposited sediments can vary largely without significant variations in grain size. The constant mean grain-size and the stable values of sorting are typical of mass-flow deposits (Mulder and Cochonat, 1996). T2 deposits showed only a thin fining-upward sequence followed by a thick homogenous interval with no inverse grading at the base. Additionally, these homogeneous layers did not present any increase of the Mn content (Fig. 2). This suggests that the turbidity currents that triggered T2 deposits did not allow oxygen to be carried to the deepest part of the basin and were thus probably related to subaquatic mass movements that reworked the sediment of the lake shores (Wilhelm et al., 2016b). A similar type of mass movement-induced facies in lacustrine environments has been previously described and called “homogenites” or “seiche deposits” (Chapron et al., 1999; Beck, 2009; Petersen et al., 2014). According to these authors, a large mass movement would induce an oscillatory movement of the whole lake-water mass. The related hyperpycnal (turbidity) currents would transport both the bed-load and suspended-load to the deepest part of the lake (Beck, 2009). The bed-load sediment would produce the coarse base of the layers, whereas the suspended-load sediment would be homogenized by the oscillation of the lake waters. Also, the remarkable reversed sequence topped by the T1b deposit (SI3) is probably related to a large subaquatic landslide due a mass movement or a very large flood, but this particular deposit is not take into account in the next sections.

5.2. Earthquake record

Earthquake shaking appears to be the most common factors that trigger mass movements in high-elevation alpine lakes (Wilhelm et al., 2016b). To test this observation, ages of most recent T2 deposits were compared with dates of regional historical earthquakes that are well documented over the last centuries (Lambert and Levret-Albaret, 1996; Scotti et al., 2004).

For this period, one T2 deposit was recorded in 1755 CE (between 1820-1710 CE at 2σ of uncertainty). To determine the earthquake that most probably triggered this deposit we follow the method provided by Wilhelm et al. (2016a,b), the epicentral MSK (Medvedev-Sponheuer-Karnik intensities) intensities of all historical earthquakes were plotted against the distance of each earthquake from the lake (Fig. SI4). The potentially recorded earthquakes

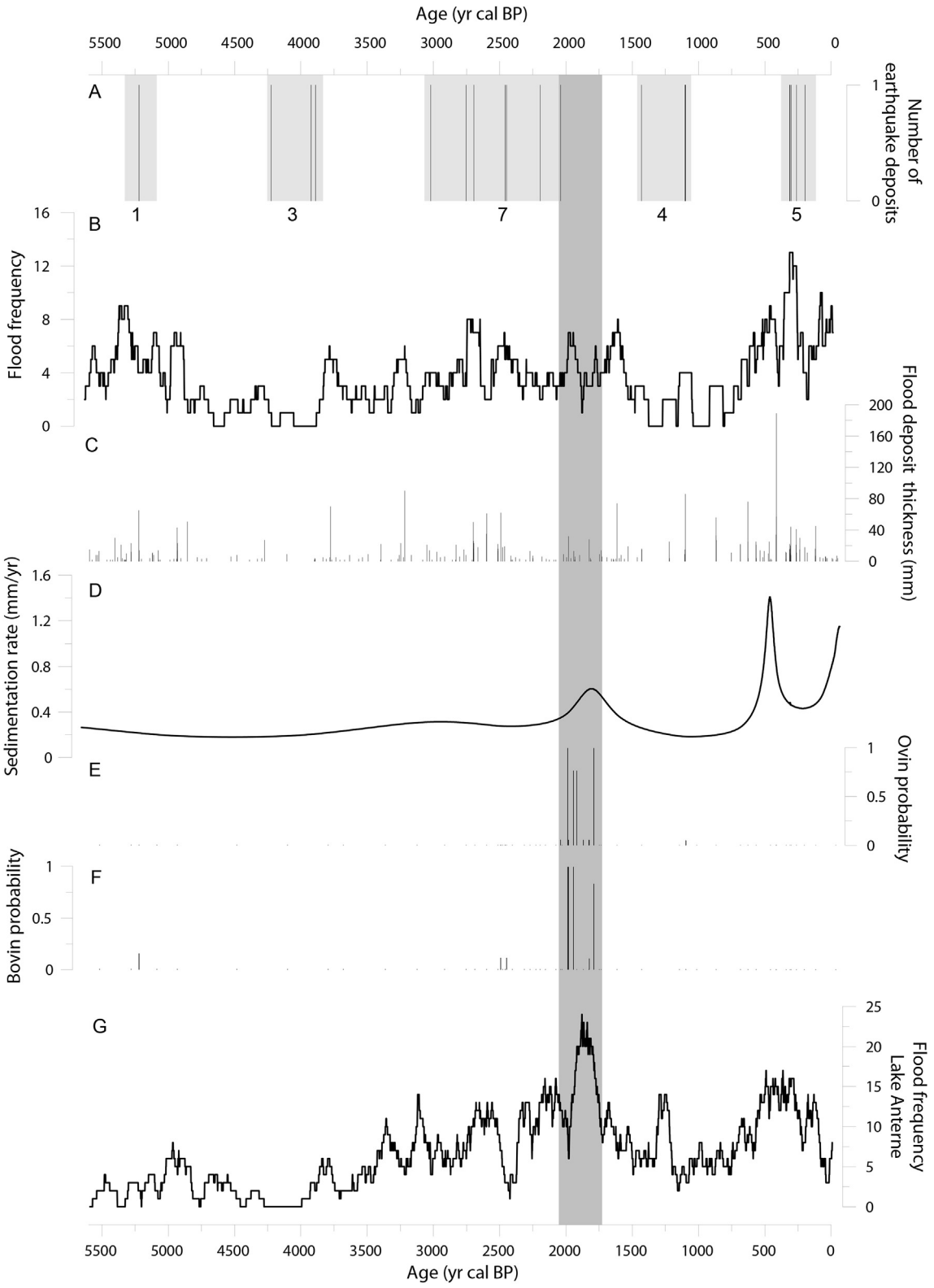


Fig. 6. Comparison of (A) reconstructed earthquake events (light grey bands represent periods of high seismic activity with the associated numbers of earthquakes during these periods), (B) flood frequency (101-yr running average) and (C) flood intensity (thickness of flood deposits), (D) average sedimentation rate without instantaneous deposits (earthquake and flood > 5 mm), (E) ovine and (F) bovine presence probability on the watershed. (G) Flood frequency of the Lake Anterne (101-yr running average) from Giguet-Covex et al. (2011); the vertical grey band is the period of high human activity in both the Savine and Anterne lake records.

are also expected to be the strongest and/or the closest to the lake, as those are expected to have generated the largest ground motions in the lake area. To take into account this second parameter, the seismic intensity of each historical earthquake in the lake area was estimated in first order by using the equation define by Wilhelm et al. (2016a,b) for this region. From this estimation, three earthquakes over the last 300 years triggered local seismic intensities above VI at 10 km from the catchment area of the Lake Savine, i.e. intensities that may be strong enough to trigger seismically induced deposits (Fig. SI4); in 1808 CE (Valle del Pellice, 8 MSK), 1785 CE (Valle del Susa, 7.5 MSK) and 1767 CE (Valle di Lanzo, 8 MSK) (Boschi et al., 2000). From this three earthquakes, the most recent and "strongest" historical earthquake occurring in 1785 CE, and could correspond to the most recent T2 deposit; dated at 1755 CE from our chronology (Fig. SI4).

A total of 20 T2 deposits supposedly induced by past earthquakes were identified. These deposits were spread in 5 clusters over the last 5.6 kyr cal BP covered by the sedimentary record (Fig. 6A). This suggests 5 periods of increased seismicity at approximately 5.22, 4.25–3.8, 3.1–2, 1.45–1.05 and 0.35–0.15 kyr cal BP. Interestingly, 10 out of 20 T2 layers presented no background sediment between adjacent deposits that can be interpreted as the record of earthquakes and respective aftershocks. The earthquake-induced deposit recorded at approximately 5.22 kyr cal BP (5.06–5.335 kyr at 2σ) can correspond to an event which was recorded at a regional scale in three other lakes located more than 115 km away in the Northern French Alps (lakes Le Bourget, Paladru and Blanc Huez; Chapron et al., 2016). This suggests the occurrence of a high magnitude earthquake.

The Earthquake Sensitivity Threshold Index, ESTI, defined by Wilhelm et al. (2016b), was estimated to be 0.15 ± 0.01 for Lake Savine, which is in the observed trend for other Alpine lakes. Note that the ESTI value was found to be highly dependent on the sedimentation rate (Wilhelm et al., 2016b). Because of high variations in sedimentation rate at Lake Savine (Fig. 5), its sensitivity to record earthquakes may have changed over time. However, despite the last cluster, these periods of relatively high earthquake occurrences mainly correspond to low sedimentation rates and could thus be associated with a regional cluster of seismicity, as it was observed for west Algeria by Ratzov et al. (2015).

5.3. Modification of erosion processes

The climatic information from the Savine flood record can be considered relevant for the last 5.6 kyr cal BP (Fig. 6B and C) if no major changes occurred in erosion processes. Erosion in the Savine catchment may be affected by geomorphological modifications and/or by land-use changes linked to human activities (i.e., grazing).

5.3.1. Geomorphological activity

Even if the modern watershed is not glaciated, substantial activity of small glaciers could have modified the sediment availability during past cold periods such as the Little Ice Age (LIA) (Koppes and Hallet, 2002; Hodder et al., 2007). However, grain-size versus thickness correlation suggests that the relationship between the intensity of flood events and the volume of transported material remained constant and was not affected by glacial cover (Fig. 3, Wilhelm et al., 2012b, 2013; Fouinat et al., 2017). Moreover, no glacial moraine is present in the watershed. Finally, the main transported sediment stream is oriented towards the south, thus limiting the effect of permafrost activity during Holocene cold periods.

5.3.2. Human activity

Lake sediment aDNA was used in this study to trace relationships between past landscape changes and past agro-pastoralism activities (mammal DNA). At this altitude, the presence of aDNA from domestic animals may be considered as a proxy for changes in grazing intensity (Etienne et al., 2013; Giguet-Covex et al., 2014; Wilhelm et al., 2016a). An increase of grazing pressure that affected both vegetal cover and soil erodibility by trampling may have made soils more vulnerable to erosion during heavy rainfalls, which led to an increasing erosion rate, and consequently, an increased sensitivity of the catchment-lake system to record flood events (Giguet-Covex et al., 2012). Mammal aDNA shows the presence of ovines and bovines in the Savine catchment only during the restricted period between 2.05 and 1.75 kyr cal BP, known as the Roman Period (Fig. 6E and F). The Roman Period with high pastoralism activity is synchronous in the Anterne lake record with the presence of both ovine and bovine herds (Giguet-Covex et al., 2014). Arnaud et al. (2016) recently noted that such an increase in human activity during the Roman Period appears to have been regional at the scale of the Northern French Alps. Indeed, the same pattern was recorded both at low (e.g., Lake Moras, Doyen et al. (2013), Lake Paladru, Simonneau et al. (2013)), intermediate (Lake La Thuile, Bajard et al., 2016) and high altitude (Lake Anterne, Giguet-Covex et al., 2011) resulting in a drastic rise in erosion fluxes for respective lakes. In Lake Anterne, Giguet-Covex et al. (2011) showed that this change of land use also resulted in the highest apparent flood frequency recorded for the Holocene (Fig. 6G) and in a distinct pattern in grain size versus flood layer thickness.

In Lake Savine, the flood frequency was relatively high during the first part of the period marked by high grazing pressure but less than during periods without grazing pressure, e.g., 5.3, 2.6, 1.7 or 0.3 kyr cal BP (Fig. 6B). Moreover, there was no change in grainsize versus event layer thickness patterns during this period (Fig. 3), as was observed at Lake Anterne because of grazing-induced changes in erosion processes (Giguet-Covex et al., 2011) or in Lake Allos in relation to human activities (Brisset et al., 2017). Therefore, it seems that the presence of ovine and bovine herds in the watershed had minor influences on the reconstructed flood chronicle.

However, the pastoralism phase corresponded to an increase in mean sedimentation rate (without flood layers) during this period. This suggests that herds and sediment sources transported during flood events were not at the same location of the watershed. The sediment source of flood events was in the southeastern part of the lake catchment with presence and probably reactivation of alluvial fans (Fig. 1), whereas the pastoralism activities were probably more restricted to the flat grassland part, i.e., close to the eastern lake shore and on the delta, as suggested by modern observations (Giguet-Covex and Walsh, pers. com.).

5.4. Palaeoflood activity

The comparison between ages of the recent flood deposits and dates of historical floods are in good agreement (Fig. 4). Moreover, significant changes in event layers related to erosion processes did not appear to be significantly linked to changes in geomorphological and/or human activities in the Savine watershed, which gives us confidence to interpret the Savine flood record from a paleoclimatic perspective.

5.4.1. Historical floods

Weather reanalysis of the last 8 floods recorded both in the Savine sediment sequence and in historical chronicles have been performed to better understand atmospheric circulation patterns (e.g., Rimbu et al., 2016). Two main groups of flood-prone synoptic configurations were identified from 500 hPa geopotential height

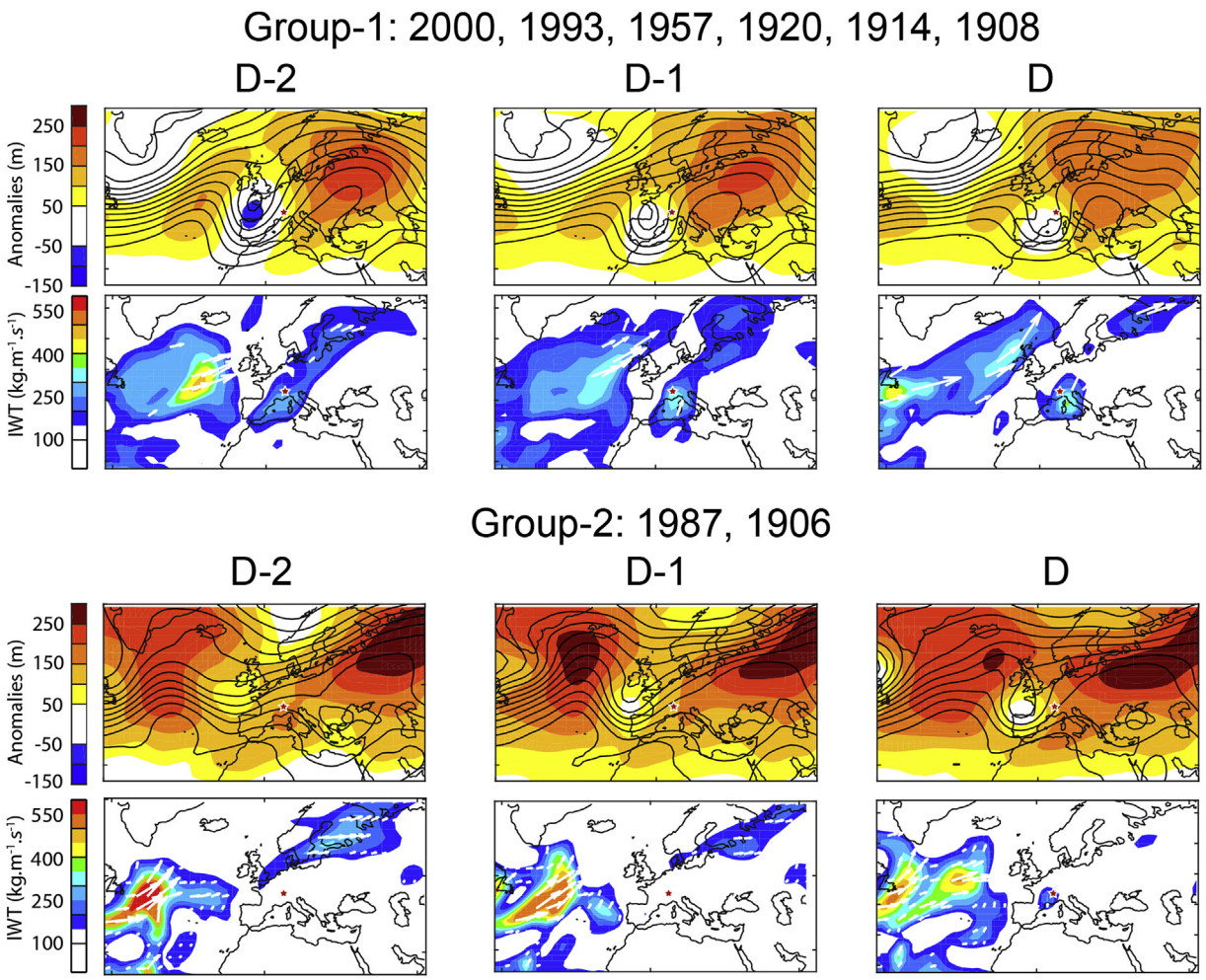


Fig. 7. Composite map of the 500 hPa geopotential heights (contour) and anomalies (shaded) and of IWT on the day of the historical flood (D) and the two days before (D-1 and D-2). Analysed events are those recorded in the sediment of Savine lake. The red star is the Savine lake location. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and IWT analysis (Fig. 7). The first group, named Group-1, corresponded to 6 of 8 flood events (15/10/2000, 24/09/1993, 12/06/1957, 24/09/1920, 23/07/1914 and 20/06/1908), and the second group, named Group-2, corresponded to 2 of 8 events (24/08/1987 and 03/07/1906). All these floods occurred between mid-June to mid-October. The 500 hPa geopotential height for Group-1 events showed a systematic low-pressure anomaly over the Bay of Biscay at D-2 that moved southeastward over the north Mediterranean area at D-1 and D. Such a pattern would trigger a cyclonic circulation around the low-pressure anomaly that would be expected to pick-up moisture from the Mediterranean Sea and favour rainfall on the Po Plain. This pattern resembles the weather pattern "WP6", identified from 1000 hPa geopotential height in Garavaglia et al. (2010), the so-called "East Return" events. The IWT for Group-1 events showed an increase of moisture in the northwestern Mediterranean area at D-2, followed by a moisture maxima and a marked advection northward into the Po Plain at D-1 and D. This atmospheric circulation pattern is typical of the late spring to autumn mesoscale events, called "East Return" events, whose circulation is associated with heavy rainfalls along the Italian border (Garavaglia et al., 2010; Gottardi et al., 2010; Lionello et al., 2012). The warm and humid air masses are vigorously uplifted by the Alpine orography, causing an abrupt cooling and intense precipitation typical of the Mediterranean climate. Two of the recorded Group-1 events (June 1957; October 2000) were regional catastrophic floods that affected many areas in the French and Italian Alps (Ratto et al., 2003; Arnaud-Fassetta and Fort, 2004). It is clear that such mesoscale precipitation events are the dominant mechanism that produces flood events recorded in the sediment sequence of Lake Savine. The geopotential height for Group-2 events shows a generalized high pressure anomaly over Europe, with two main centres located over north-eastern Eurasia and the northern Atlantic Ocean. This pattern triggers an anticyclonic circulation over the study area that hampers rainfall. It resembles the weather pattern "WP8" identified from 1000 hPa geopotential height in Garavaglia et al. (2010) that is associated with a lack of rain over Europe. The IWT for Group-2 events was null around the study site during D-2 and D-1, whereas low values of IWT were observed in a much more restricted continental area at D. The weak activity of the atmospheric circulation at the synoptic scale and the event layer occurrence during the summer season suggest that these two flood events resulted from localized phenomena such as thunderstorms.

5.4.2. *The Savine flood chronicle and climate forcing*

Because floods recorded during the last centuries were mainly triggered by mesoscale precipitation events, we assumed that the Savine palaeoflood record represented the variability of such

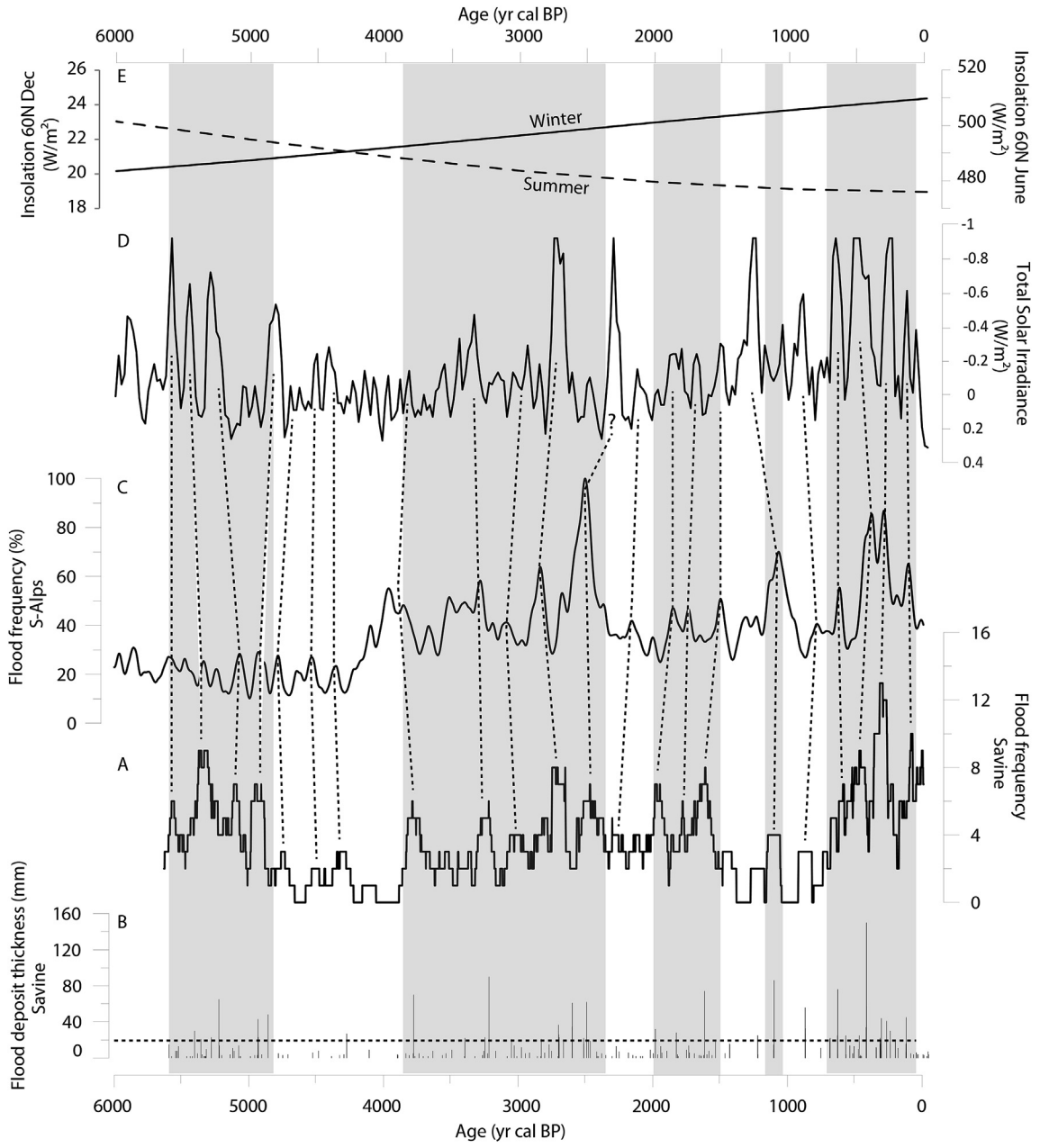


Fig. 8. Comparison of (A) reconstructed Savine flood frequency (101-yr running average) and (B) intensity (thickness of flood deposits) with (C) stacked S-Alps flood frequency (100-year low-pass filtered; Wirth et al., 2013), (D) Total Solar Irradiance (TSI) by Steinhilber et al. (2012) and (E) June and December insolation at 60°N (Berger and Loutre, 1991). Vertical grey areas mark high frequency and high intensity flood periods recorded for Lake Savine at a multi-centennial timescale. Dotted lines attempt to correlate, at a secular timescale, periods of high flood frequencies for Lake Savine with both high S-Alps flood frequencies and low solar activity periods; all these correlations are within the age model uncertainties.

events over the entire studied period (Fig. 8A and B). At a multi-centennial timescale, the Savine record showed five periods of high flood frequency between 5.6 and 4.85, 3.9–2.35, 2–1.5, 1.15–1.05 and 0.65–0.05 kyr cal BP. These increases in flood frequency were associated with the occurrence of the most intense floods (Fig. 8B). Only one intense flood occurred during a period of low flood frequency at approximately 900 yr cal BP. For the last millennium, higher flood frequency and intensity were thus recorded during the Little Ice Age (LIA, 1300 CE–1900) and lower flood frequency and intensity during the warmer Medieval Climate Anomaly (MCA, 800–1300 CE). Some events are possibly missed during cold years when the lake is covered by ice for longer than usual, such as during LIA. This suggests that there is a possibility that a flood event may occurs when the lake is still ice-bound and therefore not be recorded. However, we observed an increase of flood frequency during cold period, thus if the lake was ice covered during a longer period over a year, the interpretation does not change and could be consolidated. A similar pattern of flood frequency was observed in the European Alps (e.g., Glur et al., 2013; Swierczynski et al., 2013; Wirth et al., 2013; Wilhelm et al., 2013) and in the northwestern Mediterranean region (Moreno et al., 2008; Arnaud-Fassetta et al., 2010; Wilhelm et al., 2012a, 2016a; Benito et al., 2015). A similar pattern of flood intensity was also observed for Lake Allos, located 105 km south and strongly influenced by the Mediterranean climate (Wilhelm et al., 2012a).

Over the entire studied period, the Savine record seems similar

to the stacked flood-frequency records of the Southern Alps (Wirth et al., 2013), also strongly influenced by the Mediterranean climate (Fig. 8C). The Southern Alps record showed relatively low flood activity during the early Holocene followed by a sharp increase in flood frequency at 4.2 kyr cal BP that lasted for the entire Late Holocene. This main change at 4.2 kyr cal BP was also recorded in Lake Ledro (Vannière et al., 2013) and Lake Savine but was dated at 3.9 kyr cal BP at Lake Savine. This time lag between records may be linked to (i) the respective dating uncertainties (3.8–4.02 kyr cal BP for Lake Savine) or (ii) the more northern position of Lake Savine, which may explain a later impact of Mediterranean climate changes which propagate northward. Vannière et al. (2013) and Magny et al. (2013) attributed this transition to a nonlinear climate response to the orbital-driven gradual decrease in summer insolation at 60°N (Fig. 8E, Berger and Loutre, 1991). This would control the multi-millennial trend towards wetter conditions during the late Holocene (Mayewski et al., 2004). The Savine record also shows a generally higher level of flood frequency after 4.2 kyr cal BP, with peaks at 3.9–2.35, 2–1.5, 1.15–1.05 and 0.65–0.05 kyr cal BP. These peaks seem synchronous with the stacked S-Alps records (grey bands Fig. 8A, C) and mainly correspond to Holocene cold events (Bond et al., 2001; Mayewski et al., 2004; Wanner et al., 2011). These climate events are known to have produced a stronger meridional temperature gradient, and thus, a southward position of the westerlies with higher storm activity in the western Mediterranean region (Sabatier et al., 2012). The synchronism in flood frequency between Savine and the S-Alps records since 4.2 kyr cal BP seems still remarkable at the centennial timescale (dotted line in Fig. 8A, C). This supports the conclusion that both records were influenced by Mediterranean mesoscale precipitation events, as shown by historical events reanalysis. Before 4.2 kyr cal BP the Savine flood frequency presented a main increase between 5.6 and 4.85 kyr cal BP. This increase in flood frequency was not recorded in the Southern Alps but was visible in northern and central alpine records such as Lake Anterne (5.6–4.6 kyr cal BP, Giguet-Covex et al., 2011, Figs. 8B and 6G), Lake Bourget (centered at approximately 5.5 kyr cal BP, Arnaud et al., 2012), Ammersee (between 5.5 and 5 kyr cal BP, Czymzik et al., 2016) and in the N-Alps stack record (Wirth et al., 2013). Therefore, the flood frequency at Lake Savine seems to be in phase with the palaeoflood record of the Northern Alps before 4.2 kyr cal BP and with records of the Southern Alps after this period. Over the last millennium, floods in the Northern French Alps were mostly triggered by localized convective precipitation events during the summer (Wilhelm et al., 2013). Because floods at Lake Savine can also be triggered by thunderstorms, we propose that such processes could have been the dominant flood trigger prior to the 4.2 kyr cal BP climate transition.

The Savine flood frequency had 22 short term increases (dotted lines in Fig. 8), mostly associated with higher flood intensity. Each of these centennial increases seems to be synchronous with well-marked increases in Total Solar Irradiance (TSI, Fig. 8A, D; Steinhilber et al., 2012), which suggests an increasing flood frequency and intensity during periods of solar minima. Over the last 1.4 kyr cal BP, Wilhelm et al. (2012a) showed that major multi-centennial changes in flood frequency and intensity in the Southern Alps occurred during cooler periods such as the LIA and are associated with solar activity. Similarly, Wirth et al. (2013) suggested a strong solar influence on the Mediterranean precipitation regime because they observed a synchronism between low TSI values and high flood frequency. Similarly, in central Europe (southern Germany), Czymzik et al. (2016) found that flood frequency is higher when solar activity is reduced. Therefore, large changes in solar radiation (grand solar minima) indirectly affect climate by inducing atmospheric changes in the large European Atlantic sector (Martin-Puertas et al., 2012), which results in higher flood frequency and intensity in the Savine area. Several studies showed that periods of long-term reductions in solar radiation induced a North Atlantic storm track shift further to the south. Its maximum over the Mediterranean Sea, carried more humid air masses to central and southern Europe (e.g., Raible et al., 2007; Martin-Puertas et al., 2012). In this scenario, the part of the Alps influenced by the proximity of the Mediterranean climate likely experienced an increase in flood frequency and intensity because of increased moisture advection from the North Atlantic.

Finally, our results suggest that the occurrence and the intensity of extreme precipitation events that trigger mountain-river floods may be influenced by both lower temperatures on a millennial timescale and solar forcing through changes in atmospheric patterns on a centennial scale. Even if we cannot exclude future changes in TSI and its effect on flood occurrence, the current global temperature increases would lead to less extreme precipitation events in this area. However, higher temperatures would also lead to more efficient convective processes (Beniston et al., 2007; Giorgi et al., 2016), thereby producing heavy local rainfall events, and in turn, high intensity flood events, as they occurred prior to 4 kyr cal BP. Because this type of precipitation pattern accounts for a small part of historical events, we cannot exclude a future increase of this type of events resulting from unprecedented warming.

6. Conclusion

The high-resolution sedimentological and geochemical analyses of the sediment sequence from Lake Savine allowed the identification of 220 event layers for the last 6000 years. Twenty of the 220 event layers were differentiated by grain size parameters and geochemical features (low Mn contents) and most probably resulted from mass movements. These events were distributed in 5 clusters over this period and possibly related to earthquakes, considering their correlation to regional historic and paleo earthquakes. We found a strong relationship between deposit thickness and coarse fraction content (D90 max) for the other 200 event layers, which suggests that these event layers were produced by floods and that the thickness of deposits can be used as a high-resolution proxy for palaeoflood intensity. Local human activity was reconstructed through ancient DNA and revealed that ovines and bovines were present in the catchment area during a restricted period between 2.05 and 1.75 kyr cal BP. However, no change was observed in the grainsize versus deposit thickness pattern during this period, which suggests a minor influence of human activity on the reconstructed flood chronicle.

Weather reanalysis data for the time of historical floods recorded in the Savine sequence showed that the floods represented past variability of mainly mesoscale events, called “East Return” events (75%), and to a smaller extent, local summer convective events (i.e., thunderstorms). At a multi-centennial timescale, the Savine flood chronicle showed five periods of high flood frequency and intensity between 5.6 and 4.85, 3.9–2.35, 2–1.5, 1.15–1.05 and 0.65–0.05 kyr cal BP. These periods mainly correspond to Holocene cold events and agree well with previously reconstructed flood frequencies from the northern Alpine region between 6 and 4 kyr BP and from the southern Alpine region between 4 kyr cal BP and today. The paleohydrological transition at 4 kyr BP may have resulted from an abrupt change from Atlantic to Mediterranean climatic influences on the flood activity in the studied Alpine area and may correspond to the observed transition in the Mediterranean area during this period. At a centennial timescale, increases in flood frequency and intensity at Savine occurred during periods of solar minima. These long-term reductions in solar radiation probably indirectly affected the climate in a large European Atlantic sector through atmospheric changes, with a southward shift in the

North Atlantic storm tracks and a maximum over the Mediterranean Sea. During these periods, more humid air masses influenced by the proximity of the Mediterranean Sea were transported to the Alpine region and induced an increase in flood frequency and intensity.

Acknowledgements

This research was performed by the Hannibal Project funded by the DIPEE (Dispositifs de Partenariat en Ecologie et Environnement) Chambéry Grenoble of the CNRS INEE. We thank the municipality of Bramand for the coring authorization. The authors thank CLIMCOR Continent for coring facilities. ^{14}C analyses were acquired thanks to the CNRS-INSU ARTEMIS national radiocarbon AMS measurement programme at Laboratoire de Mesure 14C (LMC14) in the CEA Institute at Saclay (French Atomic Energy Commission). The authors thank the Laboratoire Souterrain de Modane (LSM) facilities for the gamma spectrometry measurements and Environnement, Dynamique et Territoires de Montagne for the X-ray fluorescence analyses. The authors wish to thank the editor Prof. Neil Roberts and the two anonymous reviewer who brought comments which greatly improved the original manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.quascirev.2017.06.019.

References

Alric, B., Jenny, J.P., Berthon, V., Arnaud, F., Pignol, C., Reyss, J.L., Sabatier, P., Perga, M.E., 2013. Local forcings affect lake zooplankton vulnerability and response to climate warming. Ecology 94, 2767–2780.
Appleby, P.G., 1991. 241Am dating of lake sediments. Hydrobiologia 214, 35–42.
Arnaud, F., Lignier, V., Revel, M., Desmet, M., Beck, C., Pourchet, M., Charlet, F., Trentesaux, A., Tribovillard, N., 2002. Flood and earthquake disturbance of 210Pb geochronology (Lake Anterne, NW Alps). Terra Nova 14, 225–232.
Arnaud, F., Poulenard, J., Giguet-Covex, C., Wilhelm, B., Révillon, S., Jenny, J.P., Enters, D., Bajard, M., Fouinat, L., Doyen, E., Simonneau, A., Chapron, E., Vannière, B., Sabatier, P., 2016. Erosion under climate and human pressures : an alpine lake sediment perspective. Quat. Sci. Rev. 152, 1–18.
Arnaud-Fassetta, G., Fort, M., 2004. Respective parts of hydroclimatic and anthropic factors in the recent eveolution (1956–2000) of the active channel of the Upper Guil, Queyras, Southern French Alps. Méditerranée 102, 143–156.
Arnaud-Fassetta, G., Carcaud, N., Castanet, C., Salvador, P.G., 2010. Fluviatile palaeoenvironments in archaeological context: geographical position, methodological approach and global change hydrological risk issues. Quatern. Int. 216, 93–117.
Bajard, M., Sabatier, P., David, F., Develle, A.L., Reyss, J.L., Fanget, B., Malet, E., Arnaud, D., Augustin, L., Crouzet, C., Poulenard, J., Arnaud, F., 2016. Erosion record in Lake La Thuile sediments evidences montane landscape dynamics through the Holocene. Holocene 26 (3), 350–364.
Beck, C., 2009. Late Quaternary lacustrine paleo-seismic archives in north-western Alps: examples of earthquake-origin assessment of sedimentary disturbances. Earth Sci. Rev. 96, 327–344.
Beniston, M., Stephenson, D.B., 2004. Extreme climatic events and their evolution under changing climatic conditions. Glob. Planet. Change 44, 1–9.
Beniston, M., Stephenson, D.B., Christensen, O.B., Christopher, A., Ferro, T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T., Woth, K., T, 2007. Future extreme events in European climate: an exploration of regional climate model projections. Clim. Change 81, 71–95.
Berger, A., Loutre, M.F., 1991. Insolation values for the climate of the last 10 million years. Quat. Sci. Rev. 10, 297–317.
Benito, G., Macklin, M.G., Zielhofer, C., Jones, A., Machado, M.J., 2015. Holocene flooding and climate change in the Mediterranean. Catena 130, 13–33.
Blaauw, M., 2010. Methods and code for 'classical' age-modelling of radiocarbon sequences. Quat. Geochronol. 5 (5), 512–518.
Bøe, A.-G., Dahl, S.O., Lie, Ø., Nesje, A., 2006. Holocene river floods in the upper Glomma catchment, southern Norway: a high-resolution multiproxy record from lacustrine sediments. Holocene 16, 445–455.
Bond, G., Kromer, B., Beer, J., Muscheler, R., Evans, M.N., Showers, W., Hoffmann, S., Lotti-Bond, R., Hajdas, I., Bonani, G., 2001. Persistent solar influence on north atlantic climate during the Holocene. Science 294, 2130–2136.
Boschi, E., Guidoboni, E., Ferrari, G., Mariotti, D., Valensise, G., Gasperini, P. (Eds.), 2000. Catalogue of Strong Italian Earthquakes from 461 B.C. To 1997. Introductory Texts and CD-ROM, vol. 43. Annali di Geofisica, p. 4.
Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Resour. 16, 176–182.
Brisset, E., Guitter, F., Miramont, C., Troussier, T., Sabatier, P., Poher, Y., Cartier, R., Arnaud, F., Malet, E., Anthony, E., 2017. Floods aggravated by Humans : lessons from the past. Geology. http://dx.doi.org/10.1130/G38498.1.
Campbell, C., 1998. Late Holocene lake sedimentology and climate change in southern Alberta, Canada. Quat. Res. 49, 96–101.
Calvert, S.E., Pedersen, T.F., 1993. Geochemistry of recent oxic and anoxic marine sediments: implications for the geological record. Mar. Geol. 113 (1), 67–88.
CEA, 2007. Reducing the Social and Economic Impact of Climate Change and Natural Catastrophes: Insurance Solutions and Public Private Partnerships. Insurers of Europe, Brussels.
Chapron, E., Beck, C., Pourchet, M., Deconinck, J.-F., 1999. 1822 AD earthquake-triggered homogenite in Lake Le Bourget (NW Alps). Terra Nova 11, 86–92.
Chapron, E., Simonneau, A., Ledoux, G., Arnaud, F., Lajeunesse, P., Albéric, P., 2016. French alpine foreland Holocene paleoseismicity revealed by coeval mass wasting deposits in glacial lakes (Chapter 34). In: Lamarche, G., et al. (Eds.), Submarine Mass Movements and Their Consequences, Advances in Natural and Technological Hazards Research, vol. 41. Springer International Publishing Switzerland 2016, pp. 341–349.
Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Bronnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, O., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., Worley, S.J., 2011. The twentieth century reanalysis Project. Q. J. R. Meteorol. Soc. 137, 1–28.
Corella, J.P., Benito, G., Rodriguez-Lloveras, X., Brauer, A., Valero-Garcès, B., 2014. Annually-resolved lake record of extreme hydro-meteorological events since AD 1347 in NE Iberian Peninsula. Quat. Sci. Rev. 93, 77–90.
Czymzik, M., Muscheler, R., Brauer, A., 2016. Solar modulation of flood frequency in central Europe during spring and summer on interannual to multi-centennial timescales. Clim. Past 12, 799–805.
Doyen, E., Vanniere, B., Berger, J.-F., Arnaud, F., Tachikawa, K., Bard, E., 2013. Land-use changes and environmental dynamics in the upper Rhone valley since Neolithic times inferred from sediments in Lac Moras. Holocene 23, 961–973. http://dx.doi.org/10.1177/0959683612475142.
Durant, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., Merindol, L., 2009. Reanalysis of 44 Yr of climate in the French Alps (1958–2002): methodology, model validation, climatology, and trends for air temperature and precipitation. J. Appl. Meteorol. Climatol. 48, 429–449.
Elbaz-Poulichet, F., Sabatier, P., Dezileau, L., Freydier, R., 2014. Sedimentary record of V, U, Mo and Mn in the Pierre-Blanche lagoon (Southern France) - evidence for a major anoxia event during the Roman period. Holocene 24 (10), 1384–1392.
Etienne, D., Wilhelm, B., Sabatier, P., Reyss, J.-L., Arnaud, F., 2013. Influence of sample location and livestock numbers on Sporormiella concentrations and accumulation rates in surface sediments of Lake Allos, French Alps. J. Paleolimnol. 49, 117–127.
Ficetola, G.F., Taberlet, P., Coissac, E., 2016. How to limit false positives in environmental DNA and metabarcoding? Mol. Ecol. Resour. 16, 604–607.
Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Martins Lopes, C., Boyer, F., Pompanon, F., Rayé, J., Taberlet, P., 2015. Replication levels, false presences, and the estimation of presence/absence from eDNA metabarcoding data. Mol. Ecol. Resour. 15, 543–556.
Fouinat, L., Sabatier, P., Poulenard, P., Etienne, D., Crouzet, C., Develle, A.L., Doyen, E., Malet, E., Reyss, J.L., Bonnet, R., Arnaud, F., 2017. 1700 years of interaction between glacial activity and flood frequency in proglacial Lake Muzelle (western French Alps). Quat. Res. 87 (03), 407–422.
Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garcon, R., Bernardara, P., 2010. Introducing a rainfall compound distribution model based on weather patterns sub-sampling. Hydrol. Earth Syst. Sci. 14, 951–964.
Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovicová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D., Viglione, A., 2009. A compilation of data on European flash floods. J. Hydrology 367, 70–78.
Giguet-Covex, C., Arnaud, F., Poulenard, J., Disnar, J.-R., Delhon, C., Francus, P., David, F., Enters, D., Rey, P.-J., Delannoy, J.-J., 2011. Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (Lake Anterne, 2063 m a.s.l., NW French Alps): the role of climate and human activities. Holocene 21, 651–665. http://dx.doi.org/10.1177/0959683610391320.
Giguet-Covex, C., Arnaud, F., Enters, D., Poulenard, J., Millet, L., Francus, P., David, F., Rey, P.-J., Wilhelm, B., Delannoy, J.-J., 2012. Frequency and intensity of high-altitude floods over the last 3.5ka in northwestern French Alps (Lake Anterne). Quat. Res. 77, 12–22. http://dx.doi.org/10.1016/j.yqres.2011.11.003.
Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat. Commun. 5 http://dx.doi.org/10.1038/ncomms4211.
Gilbert, R., Crookshanks, S., Hodder, K.R., Spagnol, J., Stull, R.B., 2006. The record of an extreme flood in the sediments of montane Lillooet Lake, British Columbia: implications for paleoenvironmental assessment. J. Paleolimnol. 35, 737–745.

Gilli, A., Anselmetti, F.S., Glur, L., Wirth, S.B., 2013. Lake sediments as archives of recurrence rates and intensities of past flood events. Adv. Global Change Res. In: Schneuwly-Bollschweiler, M., Stoffel, M., Rudolf-Miklau, F. (Eds.), Dating Torrential Processes on Fans and Cones – Methods and Their Application for Hazard and Risk Assessment, vol. 47, pp. 225–242.

Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., Somot, S., 2016. Enhanced summer convective rainfall at Alpine high elevations in response to climate warming. Nat. Geosci. http://dx.doi.org/10.1038/NGEO2761.

Glur, L., Wirth, S.B., Buntgen, U., Gilli, A., Haug, G.H., Schär, C., Beer, J., Anselmetti, F.S., 2013. Frequent floods in the European Alps coincide with cooler periods of the past 2500 years. Sci. Rep. 3, 2770. http://dx.doi.org/10.1038/srep02770.

Golberg, E., 1963. Geochronology with 210Pb. Radioactive Dating. International Atomic Energy Agency, Vienna, pp. 121–131.

Gottardi, F., Obled, C., Gailhard, J., Paquet, E., 2010. Statistical reanalysis of precipitation fields based on ground network data and weather patterns: application over French mountains. J. Hydrol. 432/433, 154–167.

Guédron, S., Amouroux, D., Sabatier, P., Desplanque, C., Develle, A.L., Barre, J., Feng, C., Guiter, F., Arnaud, F., Reyss, J.L., Charlet, L., 2016. A 150-year record of industrial and urban development in French Alps combining Hg accumulation rates and isotope composition in sediment archives from Lake Luitel. Chem. Geol. 431, 10–19.

Heiri, O., Lotter, A.F., Lemcke, G., 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. J. Paleolimnol. 25 (1), 101–110.

Hodder, K.R., Gilbert, R., Desloges, J.R., 2007. Glaciolacustrine varved sediment as an alpine hydroclimatic proxy. J. Paleolimnol. 38, 365–394.

IPCC, 2013. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jenny, J.P., Arnaud, F., Dorioz, J.M., Giguet Covex, C., Frossard, V., Sabatier, P., Millet, L., Reyss, J.L., Tachikawa, K., Bard, E., Pignol, C., Perga, M.E., 2013. A spatiotemporal sediment investigation highlights the dynamics of hypolimnetic hypoxia in large perialpine Lake Bourget over the last 150 years. Limnol. Oceanogr. 58, 1395–1408.

Jenny, J.P., Wilhelm, B., Arnaud, F., Sabatier, P., Giguet Covex, C., Mélo, A., Fanget, B., Malet, E., Ployon, E., Perga, M.E., 2014. A 4D sedimentological approach to reconstructing the flood frequency and intensity of the Rhône River (Lake Bourget, NW European Alps). J. Paleolimnol. 51, 469–483.

Koppes, M.N., Hallet, B., 2002. Influence of rapid glacial retreat on the rate of erosion by tidewater glaciers. Geology 30, 47–50.

Lambert, J., Levret-Albaret, A., 1996, 78. In: Catalogues d'Épicentres, Paramètres et Références, Ouest-Editions (Eds.), Mille ans de Séismes en France. Presses Académiques, Nantes.

Lapointe, F., Francus, P., Lamoureux, S.F., Saıd, M., Cuven, S., 2012. 1750 years of large rainfall events inferred from particle size at East Lake, Cape Bounty, Melville Island, Canada. J. Paleolimnol. 48, 159–173.

Lionello, P., Abrantes, F., Congedi, L., Dulac, F., Gacic, M., Gomis, D., Goodess, C., Hoff, H., Kutiel, H., Luterbacher, J., Planton, S., Reale, M., Schröder, K., Struglia, M.V., Toreti, A., Tsimplis, M., Ulbrich, U., Xoplaki, E., 2012. Introduction: mediterranean climate: background information. In: Lionello, P. (Ed.), The Climate of the Mediterranean Region, from the Past to the Future. Elsevier (Netherlands), Amsterdam. XXXV–lXXX, ISBN:9780124160422.

Lahoz-Monfort, J.J., Guillera-Arroita, G., Tingley, R., 2016. Statistical approaches to account for false positive errors in environmental DNA samples. Mol. Ecol. Resour. 16, 673–685.

Magny, M., Combourieu Nebout, N., de Beaulieu, J.L., Bout-Roumazeilles, V., Colombaroli, D., Desprat, S., Francke, A., Joannin, S., Peyron, O., Revel, M., Sadori, L., Siani, G., Sicre, M.A., Samartin, S., Simonneau, A., Tinner, W., Vannière, B., Wagner, B., Zanchetta, G., Anselmetti, F., Brugiapaglia, E., Chapron, E., Debret, M., Desmet, M., Didier, J., Essallami, L., Galop, D., Gilli, A., Haas, J.N., Kallel, N., Millet, L., Stock, A., Turon, J.L., Wirth, S., 2013. North–south palaeohydrological contrasts in the central Mediterranean during the Holocene: tentative synthesis and working hypotheses. Clim. Past. 9, 1901–1967.

Martin-Puertas, C., Matthes, K., Brauer, A., Muscheler, R., Hansen, F., Petrick, C., Aldahan, A., Possnert, G., van Geel, B., 2012. Regional atmospheric circulation shifts induced by a grand solar minimum. Nat. Geosci. 5, 397–401.

Mayewski, P.A., Rohling, E.E., Stager, J.C., Karlen, W., Maasch, K.A., Meeker, L.D., Meyerson, E.A., Gasse, F., van Kreveld, S., Holmgren, K., Lee-Thorp, J., Rosqvist, G., Rack, F., Staubwasser, M., Schneider, R.R., Steig, E.J., 2004. Holocene climate variability. Quat. Res. 62, 243–255.

Molinaroli, E., Guerzoni, S., De Falco, G., Sarretta, A., Cucco, A., Como, S., 2009. Relationships between hydrodynamic parameters and grain size in two contrasting transitional environments: the lagoons of Venice and Cabras, Italy. Sediment. Geol. 219, 196–207.

Moreno, A., Valero-Garces, B.L., Gonzalez-Samperiz, P., Rico, M., 2008. Flood response to rainfall variability during the last 2000 years inferred from the Taravilla Lake record (Central Iberian Range, Spain). J. Palaeolimnol. 40, 943–961.

Mulder, T., Cochonat, P., 1996. Classification of offshore mass movements. J. Sediment. Res. 66, 43–57.

Mulder, T., Migeon, S., Savoye, B., Faugères, J.C., 2001. Inversely graded turbidite sequences in the deep Mediterranean: a record of deposits from flood-generated turbidity currents? Geo-Mar. Lett. 21, 86–93.

Noren, A.J., Bierman, P.R., Steig, E.J., Lini, A., Southon, J., 2002. Millennial-scale storminess variability in the northeastern United States during the Holocene epoch. Nature 419, 821–824.

Pansu, J., Giguet-Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., Choler, P., 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Mol. Ecol. 24, 1485–1498.

Parris, A.S., Bierman, P.R., Noren, A.J., Prins, M.A., Lini, A., 2010. Holocene paleostorms identified by particle size signatures in lake sediments from the northeastern United States. J. Paleolimnol. 43 (1), 29–49.

Passega, R., 1964. Grain-size representation by CM patterns as a geological tool. J. Sediment. Petrol. 34 (4), 830–847.

Pedersen, T.F., Price, N.B., 1982. The geochemistry of manganese carbonate in Panama Basin sediments. Geochimica Cosmochimica Acta 46 (1), 59–68.

Petersen, J., Wilhelm, B., Revel, M., Rolland, Y., Crouzet, C., Arnaud, F., Magand, O., Chaumillon, E., 2014. Sediment archive of Lake Vens (SW European Alps, France) as a record of large magnitude earthquakes. J. Paleolimnol 51, 343–355.

R Development Core Team, 2011. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/.

Raible, C.C., Yoshimori, M., Stocker, T.F., Casty, C., 2007. Extreme midlatitude cyclones and their implications for precipitation and wind speed extremes in simulations of the Maunder Minimum versus present day conditions. Clim. Dynam. 28, 409–423.

Ratto, S., Bonetto, F., Comoglio, C., 2003. The October 2000 flooding in Valle d'Aosta (Italy): event description and land planning measures for the risk mitigation. Int. J. River Basin Manag. 1 (2), 105–116.

Ratzov, G., Cattaneo, A., Babonneau, N., Déverchère, J., Yelles, K., Bracene, R., Courboulex, F., 2015. Holocene turbidites record earthquake supercycles at a slow-rate plate boundary. Geology 43 (4), 331–334.

Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., others, 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years Cal BP.

Reyss, J.L., Schimdt, S., Legeleux, F., Bonte, P., 1995. Large low background well type detectors for measurements of environmental radioactivity. Nucl. Instrum. Methods 357, 391–397.

Richter, T.O., Van der Gaast, S.J., Koster, B., Vaars, A.J., Gieles, R., De Stigter, H.C., De Haas, H., VanWeering, T.C.E., 2006. The Avaatech XRF Core Scanner: technical description and applications to NE Atlantic sediments. In: Rothwell, R.G. (Ed.), NewTechniques in Sediment Core Analysis. : Special Publications. Geological Society, London, pp. 39–50.

Rimbu, N., Czymzik, M., Ionita, M., Lohmann, G., Brauer, A., 2016. Atmospheric circulation patterns associated with the variability of River Ammer floods: evidence from observed and proxy data. Clim. Past. 12, 377–385.

Sabatier, P., Dezileau, L., Colin, C.M., Briqueu, L., Martinez, P., Siani, G., Bouchette, F., Raynal, O., Von Grafenstein, U., 2012. 7000 years of paleostorm activity in the NW Mediterranean Sea in response to Holocene climate events. Quarternary Res. 77, 1–11.

Sabatier, P., Poulenard, J., Fanget, B., Reyss, J.L., Develle, A.L., Wilhelm, B., Ployon, E., Pignol, C., Naffrechoux, E., Dorioz, J.M., Montuelle, B., Arnaud, F., 2014. Secular interplays between pesticides and erosion in vineyards. PNAS 111 (44), 15647–15652.

Schiefer, E., Gilbert, R., Hassan, M.A., 2011. A lake sediment-based proxy of floods in the Rocky Mountain Front Ranges, Canada. J. Paleolimnol. 45, 137–149.

Scotti, O., Baumont, D., Quenet, G., Levret, A., 2004. The French macroseismic database SISFRANCE: objectives, results and perspectives. Ann. Geophys. 47 (2), 571–581.

Simonneau, A., Doyen, E., Chapron, E., Millet, L., Vannière, B., Di Giovanni, C., Bossard, N., Tachikawa, K., Bard, E., Albéric, P., Desmet, M., Roux, G., Lajeunesse, P., Berger, J.F., Arnaud, F., 2013. Holocene land-use evolution and associated soil erosion in the French Prealps inferred from Lake Paladru sediments and archaeological evidences. J. Archaeol. Sci. 40, 1636–1645. http://dx.doi.org/10.1016/j.jas.2012.12.002.

Steinhilber, F., Abreu, J.A., Beer, J., Brunner, I., Christ l, M., Fischer, H., Heikkilä, U., Kubik, P.W., Mann, M., McCracken, K., Miller, H., Miyahara, H., Oerter, H., Wilhelms, F., 2012. 9,400 years of cosmic radiation and solar activity from ice cores and tree rings. PNAS 109, 5967.

Sturm, M., Matter, A., 1978. Turbidites and varves in Lake Brienz (Switzerland): deposition of clastic detritus by density currents. Special Publ. Int. Assoc. Sedimentologists 2, 147–168.

Swierczynski, T., Lauterbach, S., Dulski, P., Delgado, J., Merz, B., Brauer, A., 2013. Mid- to late Holocene flood frequency changes in the northeastern Alps as recorded in varved sediments of Lake Mondsee (Upper Austria). Quat. Sci. Rev. 80, 78–90.

Taberlet, P., Prud'homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clement, J.C., Melodelima, C., Pompanon, F., Coissac, E., 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Mol. Ecol. 21, 1816–1820.

Vannière, B., Magny, M., Joannin, S., Simonneau, A., Wirth, S.B., Hamann, Y., Chapron, E., Gilli, A., Desmet, M., Anselmetti, F.S., 2013. Orbital changes, variation in solar activity and increased anthropogenic activities: controls on the Holocene flood frequency in the Lake Ledro area, Northern Italy. Clim. Past. 9, 1193–1209.

Wanner, H., Solomina, O., Grosjean, M., Ritz, S.P., Jetel, M., 2011. Structure and origin

of Holocene cold events. Quat. Sci. Rev. 30, 3109–3123.
Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. J. Clim. 26, 3904–3918.
Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Mon. Weather Rev. 130, 1913–1924.
Wilhelm, B., Arnaud, F., Sabatier, P., Crouzet, C., Brisset, E., Chaumillon, E., Disnar, J.R., Guiter, F., Malet, E., Reyss, J.L., Tachikawa, K., Bard, E., Delannoy, J.J., 2012a. 1.4 kyrs of flash flood events in the Southern European Alps : implications for extreme precipitation patterns and forcing over the north western Mediterranean area. Quat. Res. 78, 1–12.
Wilhelm, B., Arnaud, F., Enters, D., Allignol, F., Legaz, A., Magand, O., Revillon, S., Giguet-Covex, C., Malet, E., 2012b. Does global warming favour the occurrence of extreme floods in European Alps? First evidences from a NW Alps proglacial lake sediment record. Clim. Change. http://dx.doi.org/10.1007/s10584-011-0376-2.
Wilhelm, B., Vogel, H., Crouzet, C., Etienne, D., Anselmetti, F.S., 2016a. Frequency and intensity of palaeofloods at the interface of Atlantic and Mediterranean climate domains. Clim. Past. 12, 299–316.
Wilhelm, B., Nomade, J., Crouzet, C., Litty, C., Sabatier, P., Belle, S., Rolland, Y., Revel, M., Courboulex, F., Arnaud, F., Anselmetti, F.S., 2016b. Quantified sensitivity of lake sediments to record historic earthquakes : implications for paleoseismology. J. Geophys. Res. 121 (1), 2–16.
Wilhelm, B., Arnaud, F., Sabatier, P., Magand, O., Chapron, E., Courp, T., Tachikawa, K., Fanget, B., Malet, E., Pignol, C., Bard, E., Delannoy, J.J., 2013. Palaeoflood activity and climate change over the last 1400 years recorded by lake sediments in the north-west European Alps. J. Quat. Sci. 28, 189–199.
Wilhelm, B., Sabatier, P., Arnaud, F., 2015. Is a regional flood signal reproducible from lake sediments? Sedimentology 62, 1103–1117.
Wirth, S.B., Glur, L., Gilli, A., Anselmetti, F.S., 2013. Holocene flood frequency across the Central Alps – solar forcing and evidence for variations in North Atlantic atmospheric circulation. Quat. Sci. Rev. 80, 112–128.

Geoderma 306 (2017) 217–236

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Long-term changes in alpine pedogenetic processes: Effect of millennial agro-pastoralism activities (French-Italian Alps)☆

Manon Bajard[a,*], Jérôme Poulenard[a], Pierre Sabatier[a], David Etienne[b], Francesco Ficetola[c,d], Wentao Chen[c], Ludovic Gielly[c], Pierre Taberlet[c], Anne-Lise Develle[a], Pierre-Jérôme Rey[a], Bernard Moulin[e], Jacques-Louis de Beaulieu[f], Fabien Arnaud[a]

[a] EDYTEM, Université Savoie Mont-Blanc, CNRS Pôle Montagne, 73376 Le Bourget du Lac, France
[b] UMR INRA 42 CARRTEL, Université Savoie Mont-Blanc, 73376 Le Bourget-du-Lac, France
[c] Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble Cedex 9, France
[d] Department of Biosciences, Università degli Studi di Milano, Milan 20133, Italy
[e] Centre d'Archéologie Préhistorique du Rhône aux Alpes, F 26000 Valence, France
[f] Aix Marseille Université, CNRS, IRD, IMBE, UMR 7263 & 237, Aix-en-Provence, France

ARTICLE INFO

Handling Editor: M. Vepraskas

Keywords:
Lake sediments
Pedogenesis
Erosion
Geochemistry
Paleosols
Human activities

ABSTRACT

Human activities are known to modify soil properties; however, the associated modifications to soil processes are poorly documented, as they must be studied over long time scales. Lake Verney, which is on the Italian side of the Petit Saint-Bernard Pass in the French-Italian Alps (2188 m a.s.l.) provides a sediment record of the last 11,000 cal. yrs BP. Analysis of multiple proxies within this sediment sequence, including sedimentological characteristics, mineral geochemistry (as determined using XRF and extractable Fe fractions), pollen and non-pollen palynomorphs (NPPs) and sediment DNA (sedDNA) analysis, is compared with analyses of current soils and paleosols within this mountain ecosystem in order to understand the main drivers of long-term pedogenesis. We performed principle component analysis on both the sediment proxies and the soil geochemical properties to identify different sediment endmembers that reflect different types of soil horizons, mainly stagnic and spodic. These horizons are characteristic of specific soil processes and their associated land uses.

During the first part of the Holocene, a decrease in the carbonate fraction in the sediment reflects the loss of carbonate material from soils that occurred as early as postglacial vegetation became established. The migration of Fe-complexes until 6000 cal. yrs BP indicates the development of Podzols in the catchment. The first signs of human land use are detected at 4300 cal. yrs BP according to analyses of sedDNA and NPPs. Increases in the input of terrestrial organic matter, associated with forest clearance suggests degradation of the surface horizons of the Podzols. Erosion increased during the Roman Period due to sheep grazing. Then, while erosion was still increasing, Podzols developed into Stagnosols after the Middle Ages with cow grazing which is consistent with the current functioning of the soils. The history of the paleosols and archaeological stratigraphy within the study area confirm the model of soil evolution inferred from the lake sediments and allow us to characterize the human-induced "metapedogenetic phase" of the evolution of the soil cycle. Anthropo-pedogenesis may define the development of soils during the Anthropocene. The main consequence of this change in the functioning of the soils is a reduced sequestration of soil carbon.

1. Introduction

Soils support multiple ecosystem services that are essential for life, such as food supply and water storage or for climate regulation, such as carbon sequestration (*e.g.*, Daily et al., 1997; Adhikari and Hartemink, 2016). Because soils are a non-renewable resource over human time-scales (Vanwalleghem et al., 2017), the potential of provided services can be threatened by human activities that trigger erosion and land degradation (Oldeman, 1994; Sabatier et al., 2014; Vanwalleghem et al., 2017). In Europe, the cost of soil degradation has been estimated to be 38 billion € per year (COM EU, 2006). The worldwide loss of carbon due to land use changes corresponds to more than

☆ The data reported in this paper on the sedimentary sequence, including the age model, the loss on ignition, the grain-size distribution and the geochemistry, are deposited on PANGAEA: https://doi.pangaea.de/10.1594/PANGAEA.878058.

* Corresponding author.
E-mail address: manon.bajard@univ-smb.fr (M. Bajard).

http://dx.doi.org/10.1016/j.geoderma.2017.07.005
Received 21 April 2017; Received in revised form 30 June 2017; Accepted 9 July 2017

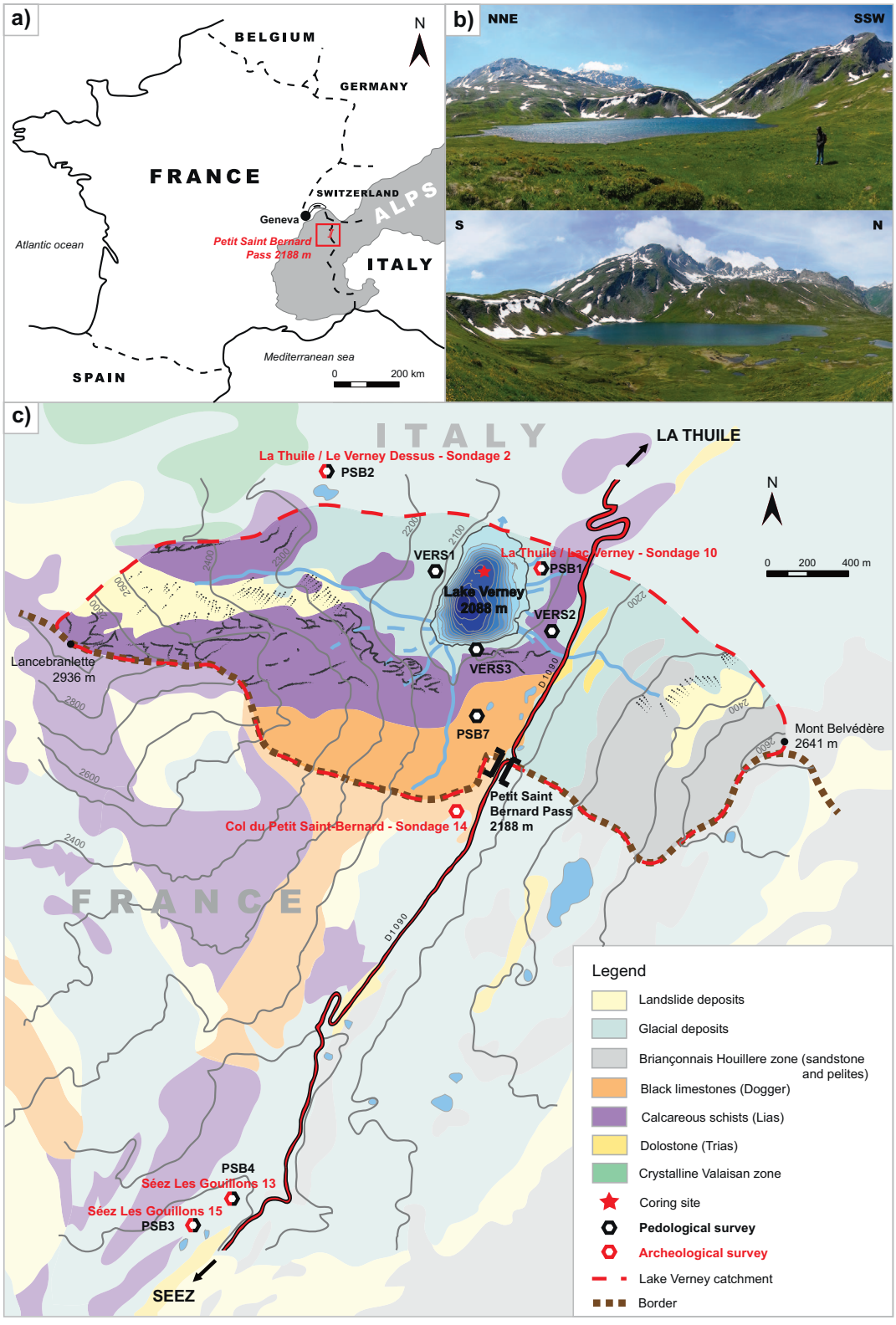


Fig. 1. Location of Petit Saint Bernard Pass (a), pictures of the Lake Verney catchment (b), locations of sampling sites and a simplified geologic map covering the area of the pass (c). Abbreviations: LTH = La Thuile.

one billion $t \cdot yr^{-1}$ since the 1960s (Le Quéré et al., 2016). However, considering the lack of knowledge concerning soil functioning, no regulatory measures have yet been established to protect soils (Puydarrieux et al., 2016). This lack is related to the complexity of pedogenesis and the feedbacks between erosion, soil properties, vegetation and agricultural management (Johnson and Watson-Stegner, 1987; Richter, 2007; Vanwalleghem et al., 2017), which have strongly varying response times.

Soils form slowly from the weathering of rocks and the incorporation of organic matter and are mixed and affected by vertical and

horizontal transfers due to hydrological fluxes, bioturbation, slopes, *etc.* Soils thus depend on soil-forming factors defined by Jenny in 1941, namely, climate, organisms, relief, parent material and time. These factors were originally described for natural soil evolution.

The role of humans in soil evolution has long been neglected (Yaalon and Yaron, 1966; Richter, 2007). However, recent studies based on paleo-records that have yielded reconstructions of both soil evolution and human activities over long-term perspectives have demonstrated the importance of the human factor in pedogenesis (Bajard et al., 2017). Metapedogenetic phases were introduced by Yaalon and Yaron (1966) to define the changes in both pedogenetic processes and soil properties that are induced by humans.

Soil erosion must be considered within a long historical context in order to better understand the current drivers of soil degradation processes and their consequences for ecosystem functioning (Trimble and Mendel, 1995; Vanwalleghem et al., 2017).

During the last few decades, several studies have noted the potential of lake sediments to record physical erosion (Giguet-Covex et al., 2011; Simonneau et al., 2013; Sabatier et al., 2014; Bajard et al., 2016; Arnaud et al., 2016), changes in the degree of soil weathering (Dearing et al., 2001; Arnaud et al., 2012; Brisset et al., 2013; Bajard et al., 2017), pedogenetic processes (Mourier, 2008; Mourier et al., 2008; Giguet-Covex et al., 2011; Malkiewicz et al., 2016; Hošek et al., 2017), and the progressive and regressive phases of pedogenesis (Brisset et al., 2013; Malkiewicz et al., 2016; Bajard et al., 2017). Lakes sediments indeed store soil particles obtained from catchment fluxes and erosion triggered by human activities and/or climatic changes that correspond to soil development at the date of storage, providing a record of their evolution. Changes induced by erosion will depend on the characteristics of the soil profile, and thus into early pedogenetic processes. Few studies on soil evolution deciphered from lake sediments succeed in making accurate observations during the Anthropocene (Mourier et al., 2008; Giguet-Covex et al., 2011; Malkiewicz et al., 2016) because, among other factors, catchments drain the soil landscape diversity (Dreibrodt and Wiethold, 2015). It is necessary to understand the interdependence of landscape history (*i.e.*, *vegetation changes, human activities*), soil evolution and erosion to sort the main forcing factors and their consequences on the agro-ecosystem to adapt its future management and ensure its sustainability.

Located in the Northwestern Alps, the Petit Saint-Bernard Pass area marks currently the border between France and Italy. Archaeological surveys and paleo-environmental studies enable the reconstruction of the landscape evolution and occupation phases around this pass. The oldest occupation phases were evident since the Neolithic period, and occurred mainly during the early Bronze period (*ca.* 4000 cal. yrs BP), the late Bronze period (3000 cal. yrs BP) and then in the Roman and Medieval periods (Crogiez-Petrequin, 2006; Leveau and Segard, 2006; Miras et al., 2006; Talon, 2006; Rey et al., 2014). Paleosols and archaeological stratigraphy indicated different occupation phases and recurrences within different soil horizons. Thus, we have an ideal site where human activities are well-documented, to assess their effects on soil evolution. Pollen and subfossil chironomid analysis from seven peat bog records collected around the pass at elevations ranging from 1850 to 2345 m a.s.l., indicate that changes in the landscape and climate occurred throughout the Holocene (Millet et al., 2008; Miras et al., 2006), but do not permit characterization of the effects of human activities on soil erosion dynamics. In this context, we studied the sediments from the largest lake on the Italian side, Lake Verney, to reconstruct, past soil evolution and human activities throughout the Holocene, using multiproxy lake sediment archives and present-day soil analysis. Data derived from paleosols (Moulin and Rey, 2008; Rey et al., 2014), charcoals (Talon, 2006) and archaeology (*e.g.*, Rey and Moulin, 2006; Rey et al., 2014) then allowed us to validate and complete the Lake Verney sediment record and provide a complete soil evolution at the landscape scale. Finally, we integrated this evolution into the evolution cycle of the ecosystem.

2. Materials and methods

2.1. Study site

The Petit Saint Bernard Pass (2188 m a.s.l.) is located in the Western Alps, on the French-Italian border (Fig. 1-a), and it separates the villages of Séez in the Tarentaise valley (France) and La Thuile in the Aoste valley (Italy). On the Italian side of the pass, Lake Verney (2088 m a.s.l.) has a surface of 0.2 km^2 and a maximum depth of 21 m. The lake is fed by three main streams that flow from the south. Its catchment covers an area of 3.5 km^2 and rises up to 2900 m a.s.l. According to the weather observation network of the alpine massif (ROMMA), the mean temperatures at the pass range from − 3.4 °C in winter to 9.7 °C in summer (these values represent averages for 2014–2016). Snow covers the area from November to June, and the lake is ice-covered during winter. The western part of the catchment is underlain by Jurassic calcareous schists (Lias) and black limestones (Dogger). The eastern side of the pass is characterized by sandstone and pelites of the Briançonnais Houillere zone (Fig. 1-c). There are currently no trees in the catchment. Around the lake, grazed grasslands are dominated by matgrass and thufurs and are punctuated with *Vaccinium myrtillus* and *V. uliginosum* and several wetlands (Fig. 1-b). The current upper tree line (mainly *Picea abies*) is situated at approximately 1950 m a.s.l. on both sides of the pass. Anthracological analysis indicated an upper tree line of *Pinus cembra* between 2300 and 2500 m before human clearing (Talon, 2006). Fire episodes were recorded regularly in soils, suggesting that pasture activities occurred continuously until the present. Several remains of pastures structures were identified, mainly for cattle farming, in addition to Roman buildings situated along the Roman road that crossed the pass (Leveau and Segard, 2006; Rey et al., 2014). Surveys and pedo-sedimentary profiles performed on a 250-km^2 area covering both sides of the pass for archaeological analysis showed specific soil patterns above 1900 m, with mainly Stagnosols and sparse Podzols (Moulin and Rey, 2008; Rey et al., 2014).

2.2. Soil sampling

An archaeological and pedo-sedimentary survey was performed between 2003 and 2007 in the area of the Petit Saint-Bernard Pass (Moulin and Rey, 2008; Rey et al., 2014) within the framework of the Interreg III A 2003–2006 project. We used anew the 5 profiles from this earlier study that are closest to the pass (LTH/Lac Verney – Sondage 10, LTH/Le Verney Dessus – Sondage 2, Séez – Les Gouillons 13, Séez - Les Gouillons 15 and PSB7; the names of these profiles are simplified to PSB1, PSB2, PSB4 and PSB3, respectively) for soil analyses (Fig. 1-c). We sampled three new profiles on each side of the lake: VERS1, VERS2 and VERS3. PSB1 (2100 m a.s.l.) is located north of Lake Verney. PSB2 (2262 m a.s.l.) is located near Lake Verney Dessus, in the northern part of the Lake Verney catchment. VERS 1 and VERS 2 are located on either side of the lake; VERS3 is to the south of the lake. PSB7 is located on the flat area west of the pass on the Italian side. PSB4 and PSB3 (2000 m a.s.l.) are situated on the French side of the pass. Soil profiles were described with color, texture, structure and HCl effervescence test. The FAO (Food and Agriculture Organization) soil classification (WRB - FAO, 2014) and the Guidelines for soil description (FAO, 2006) were used for horizon and soil denominations. At least one sample was collected per horizons and sieved to 2 mm for further laboratory analyses.

2.3. Sediment sequence and chronology

2.3.1. Coring

Two coring missions were undertaken to retrieve sediments from the lake Verney in autumn 2012 and winter 2013. Short gravity cores were taken in 2012 from a small boat, and three longer sections (VER13-01, 02 and 03) that are approximately 2.4 m each in length

were retrieved in 2013 from the ice-covered surface of northern flat part of the lake (45°41′19.8″N, 6°52′57.06″E) at a depth of 15 m using an Uwitec piston coring device. The deepest part of the lake was avoided regard to rock landslides south of the lake. The cores were split into two halves. Each half was described in detail, and pictures were taken with a 20-pixel·mm^{-1} resolution. The lithological description of the cores allowed the identification of different sedimentary units. A composite profile VER13-I was built using XRF-correlation from the three sediment sections obtained during the 2013 coring mission (the IGSN codes IEFRA05GT, IEFRA05GU and IEFRA05GV correspond to VER13-01, VER13-02 and VER13-03, respectively; *IGSN codes refer to an open international database*, *www.geosamples.org*). Data from a short gravity core (VER12-P4) obtained in 2012 (IGSN code: IEFRA00BU) were included to provide a well-preserved water-sediment interface.

2.3.2. XRF core scanner

The relative contents of major elements were analyzed by X-ray fluorescence (XRF) at high resolution (1 mm sampling step) on the surface of the sediment core with an Avaatech Core Scanner (EDYTEM Laboratory, CNRS-Université Savoie Mont-Blanc). The X-ray beam was generated with a rhodium anode and a 125-µm beryllium window, which allows a voltage range of 7–50 kV and a current range of 0–2 mA. The split core surface was first covered with a 4-µm-thick Ultralene film to avoid contamination and desiccation of the sediment. The element intensities were expressed in counts per second (cps). The geochemical data were obtained with different settings according to the elements analyzed. These settings were adjusted to 10 kV and 2 mA for 20 s to detect Si, Ca, Al, Fe, Ti, K, Mn, and S. For heavier elements (*i.e.*, Sr, Rb, Zr, Br, and Pb), measurements were performed at 30 kV and 1.2 mA for 80 s.

2.3.3. Grain-size distribution

The grain-size distribution of the sediment was determined at high resolution, every 1 cm, using a Malvern Mastersizer 2000G laser particle sizer. Ultrasonics were used to dissociate mineral particles and to avoid their flocculation. Results of the grain-size distribution were processed with MATLAB software and presented as a contour plot with a color scale according to the percentage abundance of particles for each grain-size class.

2.3.4. Sedimentary environmental DNA (sedDNA)

To reconstruct past vegetation and the nature of herds (*i.e.*, cows or sheep) by DNA analysis, we sampled 55 slices of 2.5 to 3-cm thickness covering the whole sediment core, following the strict laboratory precautions described in Giguet-Covex et al. (2014). For each sediment slice, we mixed 15 g of sediment with 15 ml of saturated phosphate buffer (Na2HPO4; 0.12 M; pH ≈ 8) for 15 min. The mixture was then centrifuged (10 min at 10,000 *g*). 12 ml of the resulting supernatant were transferred to Amicon® Ultra-15 10 K Centrifugal Filter Devices (Millipore) and centrifuged (20 min at 4000 *g*) for sedDNA concentration. 400 µl of the resulting concentrate was kept as starting material for DNA extraction using the NucleoSpin® Soil kit (Macherey-Nagel) (Taberlet et al., 2012). Four extraction controls were performed. Mammal and plant DNA was amplified with the primer pair P007, following the protocol described in Giguet-Covex et al. (2014). In addition to extraction controls, we also performed three PCR controls, containing PCR mix but no DNA template. Four PCR positive controls were also performed, each of which contained 0.18 ng of DNA extracted from a marsupial (*Didelphis marsupialis*), that is not found in Europe. All samples and controls were amplified in 12 replicated PCRs (Ficetola et al., 2015). Sequencing was performed by 2 × 125 bp pair-end sequencing on an Illumina HiSeq 2500 platform, which returned 21,192,000 reads.

DNA sequences were filtered using the OBITOOLS software (Boyer et al., 2016), as described in Pansu et al. (2015). Sequences were assigned to the relevant taxa by comparing them with global databases of mammal and plant sequences that were generated from EMBL using the ecoTag program. For mammals, we only kept sequences with a match > 97% with a mammal genus. For plants, we focused on two species characteristic of the study area (pine: *Pinus* sp. and spruce: *Picea* sp.), and we only considered sequences with a match > 97% with the sequences of these species. The full set of sedDNA results will be published in a subsequent study (Chen et al. in prep.). In sedDNA studies, false detections (*i.e.*, sporadic detections of absent species) are possible. We thus only considered species confirmed by multiple PCR analyses performed on the same sample, whereas we discarded sequences detected with < 5 reads in a given PCR replicate, and sequences frequently detected in the controls (human and pig).

2.3.5. Pollen and non-pollen palynomorphs (NPPs)

Following the same sampling that of sedDNA, 51 subsamples of approximately 1 cm^3 were prepared for pollens and non-pollens palynomorphs analysis, using the standard procedure described by Faegri and Iversen (1989). *Lycopodium clavatum* tablets were added to each subsample (Stockmarr, 1971) in order to calculate the concentration of NPPs. They were counted following the procedure described by Etienne and Jouffroy-Bapicot (2014) and are expressed as accumulation rates (no·cm^{-2}·yr^{-1}). Their ecological interpretation in term of agro-pastoral activities and OM origin follows the work of Doyen and Etienne (2017). Non-arboreal pollen (NAP) was expressed as a percentage of the total pollen.

2.3.6. Chronology

The chronology of the Lake Verney sediment sequence is based on nine ^{14}C measurements performed on terrestrial plant macroremains, as well as short-lived radionuclide measurements (^{210}Pb and ^{137}Cs) to date the latest century. AMS radiocarbon dates were performed by accelerator mass spectrometer (AMS) at the Poznan Radiocarbon Laboratory and at the Laboratoire de Mesure 14C (LMC14) ARTEMIS at the CEA (Atomic Energy Commission) Institute of Saclay (Table 1). The ^{14}C ages were converted to 'calendar' years using the IntCal13 calibration curve (Reimer et al., 2013). Short-lived radionuclide were measured on the

Table 1
Radiocarbon ages for Lake Verney sediment sequence.

Sample name	Lab. code	Depth (mm)	Sample type	Age (BP)	Min. age (cal. BP)	Max. age (cal. BP)
VERNEY1	**Poz-73372**	**393**	**Wood and twigs**	**8000 ± 50**	**8656**	**9010**
VER13-1	SacA 38343	548	Plant macroremains	875 ± 30	726	907
VERNEY2	**Poz-73373**	**753**	**Herbs, twigs and moss**	**4615 ± 35**	**5145**	**5465**
VER13-02A-88	SacA 32338	1048	Plant macroremains	2100 ± 35	1989	2287
VERNEY3	**Poz-73374**	**1213**	**Herbs, twigs and moss**	**3350 ± 35**	**3481**	**3687**
VERNEY4	Poz-73375	1353	Herbs, twigs and moss	3940 ± 50	4241	4521
VER13-02B-61	SacA 32339	1748	Plant macroremains	8890 ± 45	9793	10,188
VER13-02B-96	SacA 32340	2108	Plant macroremains	9725 ± 50	10,876	11,240
VER13-3	SacA 38345	2383	Plant macroremains	9780 ± 60	10,905	11,315

Bold samples correspond to excluded dates from age-depth modeling.

top of the sediment sequence at the Modane Underground Laboratory (LSM) by gamma spectrometry, using high-efficiency, very low-background, well-type Ge detectors (Reyss et al., 1995). The isotope ^{137}Cs was accidentally introduced into the environment at the end of the 1950s as by-product of atmospheric nuclear weapons tests (peak at AD 1963). The Chernobyl accident in 1986 also dispersed ^{137}Cs into the northern atmosphere (Appleby et al., 1991). ^{210}Pb excess ($^{210}Pb_{ex}$) results in the disintegration of ^{226}Ra of rocks into ^{222}Rn which disintegrates, in gas form, in $^{210}Pb_{ex}$ in the atmosphere and deposited in the lake. It was calculated in the sediment as the difference between total ^{210}Pb (mixing of erosion and atmospheric deposits) and ^{226}Ra activity (Goldberg, 1963). ^{226}Ra activity is assuming equivalent to ^{210}Pb activity of sediments triggered by erosion. Then, we used the Constant Flux/Constant Sedimentation (CFCS) model and the decrease in $^{210}Pb_{ex}$ to calculate the sedimentation rates. The uncertainty of sedimentation rates obtained by this method was derived from the standard error of the linear regression of the CFCS model. The age–depth model was then generated using R software and the R code package 'Clam' version 2.2 (Blaauw, 2010).

2.4. Soil and sediment analysis

2.4.1. Loss on ignition

Loss on ignition (LOI) analysis was performed on both the soil samples and sediments, following a continuous 1 cm sampling step all along the sediment sequence and every 3 cm using the sampling of DNA, pollens and NPPs analysis to estimate organic matter (OM) and carbonate content following the procedure described in Heiri et al. (2001). OM content was measured on dried sediment after 4 h at 550 °C in an oven, then carbonate content was measured after 2 h at 950 °C.

2.4.2. P-XRF mineral geochemistry

A portable ED-XRF spectrometer (S1 TITAN Bruker) was used to measure the major elements, which were expressed as relative percentages of oxides. Analysis of both gently crushed soil and sediment samples was performed through a 4-μm-thick Ultralene film in a 32-mm-diameter plastic cup. The samples were triplicated and analyzed over 60 s using the GeoChem Standard instrument's internal calibration mode (Shand and Wendler, 2014). The organic matter content that was deduced from the LOI at 550 °C was added to the sum of all major and trace elements. Consequently, the elemental results were brought to 100% less the OM percent to overcome the closed sum effects that are linked to OM variations (Baize, 2000). The standard deviations of the replicates were lower than the instrument errors, which were thus conserved as measurement uncertainties. The total contents of Fe and Al (Fe_{Tot} and Al_{Tot}) were estimated from Fe_2O_3 and Al_2O_3 contents, respectively.

2.4.3. Extractable fractions

To identify pedogenetic compounds specific to Podzol spodic horizons in both the soil and sediment samples, we quantified the amount of secondary mineral phases containing Fe, Al and Si. Secondary Fe-bearing phases were extracted from PSB1, PSB2, PSB3, PSB4 and PSB7 and secondary Fe-, Al- and Si-bearing were extracted from VERS1 and VERS2, and from 38 samples (each 1 cm thick) taken from the sediment core. For PSB1, PSB2, PSB3, PSB4 and PSB 7, the secondary Fe-bearing phase were extracted with three solutions of dithionite citrate, acid ammonium oxalate and 0.1 M Na pyrophosphate to quantify the crystalline, amorphous and organic-bound fractions in the samples, respectively (McKeague et al., 1971). For VERS1 and VERS2 and the sediment samples, Al, Fe and Si were extracted only using an acid ammonium oxalate solution (Tamm, 1922). All of the extracts were then analyzed by ICP-AES. Soil pH was measured with a 1:5 soil-water ratio.

2.5. Statistical analysis

Principal component analysis (PCA) was performed on two sedimentary datasets that corresponded to the sampling strategy used for sedDNA, pollen and NPPs (3-cm-thick samples) and to the sampling strategy used for selective extraction (1-cm-thick-samples), as well as on the soil geochemical parameters. PCA was used to assess the relationships between the elemental variations, human activities and soil dynamics in the catchment (Sabatier et al., 2010). Analyses were performed using R 2.13.1 (R Development Core Team, 2011).

3. Results

3.1. Soil characterization

Apart from Cambisols, Leptosols, Regosols and sparse Podzols, the current soils of the Petit-Saint Bernard Pass area, at elevations between 1800 and 2700 m are mainly related to Stagnosols (Moulin and Rey, 2008).

Regosols, represented by VERS3, were identified in the south part of Lake Verney on black limestone (Fig. 1-c). They are thin, and only the coarse elements effervesce with HCl acid solution. pH is high (pH = 7.1). No horizons were distinguished in these soils but they received sediments from slope deposits. Thus, they may also develop on the upper slopes of the catchment, locally as Leptosols and Cambisols.

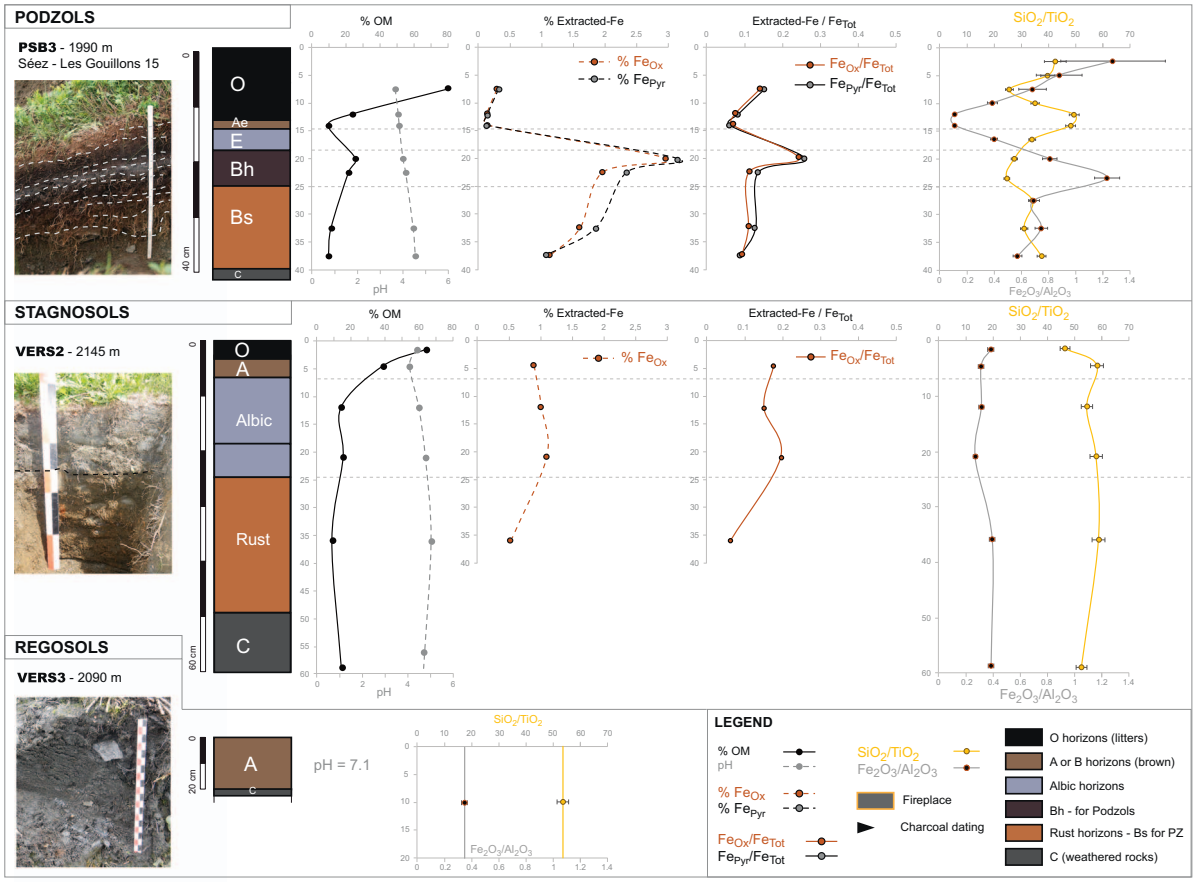
Stagnosols were found around the lake (VERS1, VERS2), within the pass (PSB7 but also Séez - Col du Petit Saint-Bernard - Sondage 14) and on the Italian side of the pass below the lake catchment near Plan Veyle at an elevation of 2000 m (Rey et al., 2014). They are characterized by low pH (4.5) and low organic matter content in the mineral horizons (Fig. 2-a). Two types of mineral horizons, albic horizons in the subsurface and rust-colored horizons below, were noted in these soils (Fig. 2).

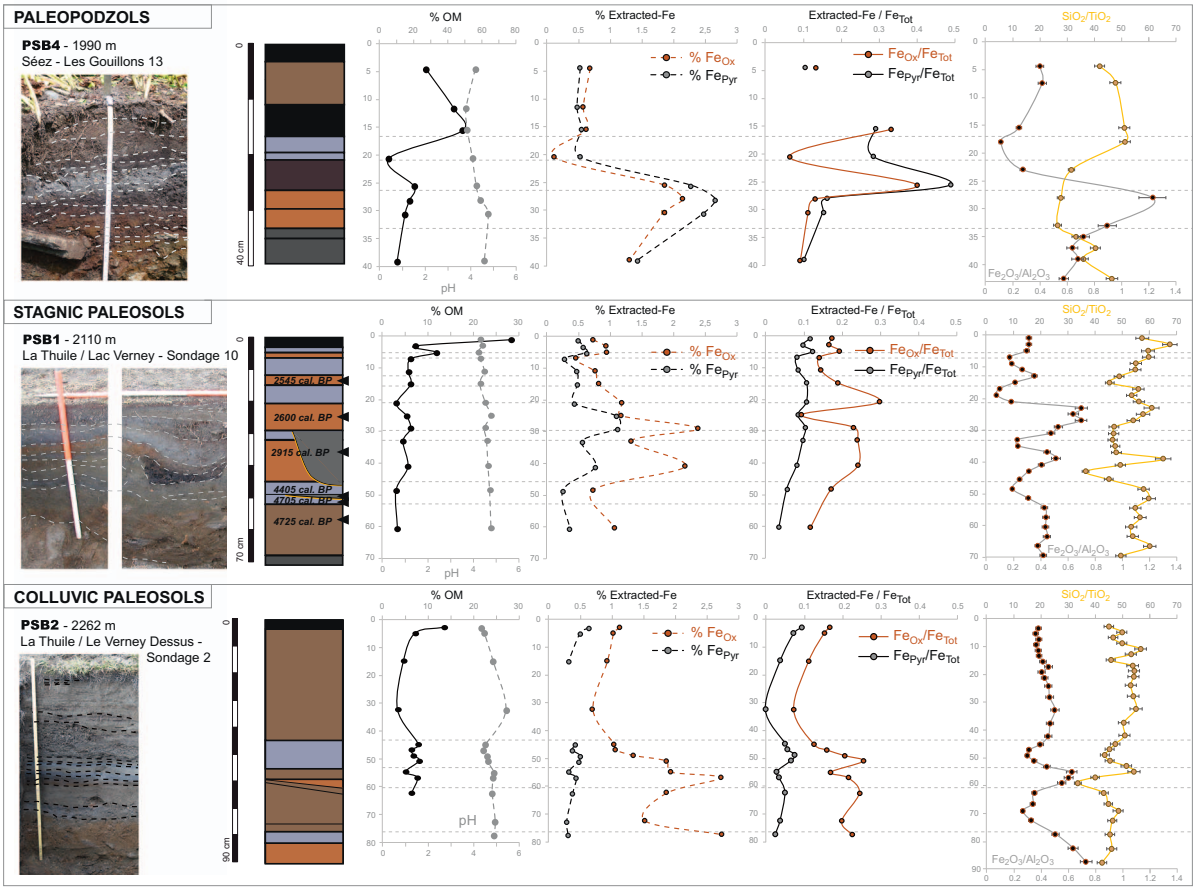
Podzols were found on the French side of the pass and are characterized by thick surface organic (O) horizon, with light gray (E) and rust (Bh and Bs) horizons below (Fig. 2-a, *e.g.*, PSB3). Organic matter is also more important in the Bh horizons than in the eluvial E horizon. The pH values of these soils are also acid, in the same range of that of Stagnosols, between 3.7 and 4.8 (Fig. 2).

The amount of Fe extracted using ammonium oxalate acid and 0.1 Na-pyrophosphate solutions were < 2% in the Stagnosols and exceeded 2% in the spodic horizons of the Podzols, whereas the eluvial horizons showed very low (< 0.15%) amount of extracted Fe. Both Fe_{Ox}/Fe_{Tot} and Fe_{Pyr}/Fe_{Tot} ratios in Podzols follow the same trend: iron depletion in the eluvial horizons (ratios < 0.1) and enrichment in the spodic horizons (> 0.25). The Fe_2O_3/Al_2O_3 ratio presents the same variations, but we note that there is more Al_2O_3 than Fe_2O_3 is present in the upper part of the spodic horizon (Fe_2O_3/Al_2O_3 = 0.8) than in it lower part (Fe_2O_3/Al_2O_3 = 1.2). In contrast to the pattern displayed by Fe, the SiO_2/TiO_2 ratio shows an increase in the eluvial horizon (SiO_2/TiO_2 = 50) compared to spodic horizons (SiO_2/TiO_2 = 25). In the Stagnosols, the ratios do not present such pronounced variation as those seen in the Podzols. The values of Fe_{Ox}/Fe_{Tot} and Fe_{Pyr}/Fe_{Tot} ratios are < 0.2, the values of Fe_2O_3/Al_2O_3 are largely constant and range between 0.2 and 0.4, and the values of the SiO_2/TiO_2 ratio are higher, ranging between 40 and 60.

Stagnosols are driven by hydrological processes and clogging of the surface (albic) horizons, especially during snowmelt, leading to the migration of Fe (II) and forming rust horizons below. Podzols are the result of the formation of organo-metallic compounds of Fe and Al in the surface horizons (O and E) and their migration with acidification from the eluvial horizon to the spodic Bh and Bs horizons.

Three paleosols were identified, in PSB1, PSB2 and PSB4 (Fig. 2-b). They correspond to old soils that are covered by deposits that are now developing new active soil layers. PSB1 and PSB2 display recurrences of albic and rust horizons (Fig. 2-b). The lowest rust horizons have high





(caption on next page)

Fig. 2. Analyses of soils in the Petit Saint Bernard Pass area according to present (A) and past (B) pedogenesis. Organic matter (OM) contents, pH, Fe contents extracted using oxalate and pyrophosphate solutions, normalized ratios of extracted Fe to total Fe, and SiO_2/TiO_2 and Fe_2O_3/Al_2O_3 ratios are shown according to depth. Photos and descriptions by P-J Rey and B. Moulin.

contents of Fe_{Ox} and display variations in Fe_2O_3/Al_2O_3 and SiO_2/TiO_2 ratios, similar to those seen in the Podzols. PSB2 was covered by a succession of coarse colluvial deposits over approximately 40 cm in thickness. PSB1 includes two hearths at depth of 40 and 50 cm that have been dated using *Pinus* charcoals to 2915 and 4405 cal. yrs BP, respectively (Rey et al., 2014). Four other levels occurring at depths of 15, 25, 52, and 58 cm within this profile were dated using *Pinus* charcoals to 2545, 2600, 4705 and 4725 cal. yrs BP. PSB4 presents an underlying succession of horizons similar to the PSB3 Podzols. Podzolic patterns were confirmed by the analysis of OM and extracted Fe. A Cambisol had developed on this paleopodzol.

3.2. Statistical analyses of the soil horizons

Fe_{Pyr}, which represents the content of Fe-organic complexes, was not measured in the horizons of VERS1 and VERS2, but as Si_{Ox} was < 0.01% in these horizons (data not shown), we considered that the abundances of amorphous compounds were negligible. Thus, we assumed that Fe_{Ox} would be equal to Fe_{Pyr} in VERS1 and VERS2 and carried out the PCA using a Fe_{Pyr}/Fe_2O_3 ratio for all the horizons.

The correlation circle from the PCA and its corresponding biplot map for the soil samples highlighted the relations between the different elements and ratios measured in soils horizons, and allowed us to characterize each type of horizon (Fig. 3-a). Dimensions 1 and 2 (denoted as Dim 1 and Dim 2) represent 72% of the total variability. Several end members were identified on the correlation circle. The first one, which is dominated by SiO_2 and the K_2O/TiO_2 and Al_2O_3/TiO_2 ratios, is positively correlated with the first dimension of the PCA and opposite to Fe_{Pyr}/Fe_2O_3 that is poorly and negatively correlated with this first component. A second end members is positively correlated with the second component of the PCA and yields high positive loadings for TiO_2, % Fe_{Ox}, Fe_2O_3 and Fe_2O_3/TiO_2. It adjoins Fe_{Pyr} and Fe_2O_3/Al_2O_3 which are negatively correlated with Dim 1 and positively correlated with Dim 2. On the other hand, the SiO_2/TiO_2 ratio is positively correlated with Dim 1 and negatively correlated with Dim 2. Moreover, the organic matter and CaO contents are negatively correlated with both components and anticorrelated with the K_2O and Al_2O_3 contents. Mapping of the horizons highlights the correlation between the surface O horizons and organic matter, and between the bleached (albic and E) horizons with SiO_2/TiO_2, but there is no possibility of distinguishing the eluvial horizons of the Podzols from the albic horizons of Stagnosols (Fig. 3-a). In contrast, this PCA allows us to distinguish spodic (Bs and Bh) horizons of the Podzols, which are highly correlated with % Fe_{Pyr} and Fe_2O_3/Al_2O_3, from the rust horizons of the Stagnosols, which are positively correlated with both components of the PCA that are represented by Al_2O_3 and K_2O contents. The A and B horizons, which are poorly differentiated, are located in the middle of the map and are not associated with any component. The C horizons are characterized by poorly weathered material and are mainly related to the second component of the PCA, which is associated with the TiO_2 and Fe_2O_3 contents. We note that the surface horizons are mainly negatively correlated with the second component, and the deeper horizons are positively correlated with this component. Apart from the E horizons, the first component permitted the separation of horizons experiencing podzolization which display positively correlated OM enrichment and spodic properties, from horizons with stagnic properties, where these characteristics are negatively correlated.

3.3. Lithology and sedimentological analyses

Based on macro-observations of the sediment surface and on the sedimentary and geochemical analyses presented in Fig. 4, four sedimentary units were identified.

Unit 4 (2.4 to 2.12 m) is composed of green-gray silts to sands and pebbles. The LOI at 550 °C presents the largest values of the sediment sequence. The grain-size distribution differs strongly from the rest of the sequence and is coarser with particles ranging in size from 40 to 200 μm (Fig. 4). The XRF-intensities of Ti and the contents of SiO_2 and Al_2O_3 present the lowest values of the sequence. The Ca intensities and Fe_2O_3 are variable whereas the SiO_2/TiO_2 ratio remains constant.

Unit 3 (2.12 to 1.7 m depth) is made of dark gray silt and includes some orange layer that alternate with layers containing many macrofossils and sparse fine sandy layer (Fig. 4). Unit 3 includes the largest value of LOI at 950 °C and Ca intensities (approximately 25%). The grain size displays a slight fining-upward trend. All of the other geochemical parameters are highly variable, except for the SiO_2/TiO_2 ratio that seems constant but is higher than in Unit 4.

Unit 2 (1.7 to 1.15 m depth) is composed of gray silt with alternating fine sand and orange patterns (Fig. 4). Both the Ca intensities and LOI at 550 °C decrease with time. The other parameters increase with time, except for the Al_{Ox}/Al_{Tot} and SiO_2/TiO_2 ratio which remain constant; the latter ratio has higher values than the previous unit, and the Fe_2O_3/Al_2O_3 ratio decreases in the upper part of the unit. The grain size does not show important variations.

Unit 1, which extends from a depth of 1.15 m to the top, is made of dark gray to brown silts and includes fine sand layer alternating with orange patterns (Fig. 4). The 550 °C LOI is quite constant in this unit, approximately 5%, and the LOI 950 °C varies slightly at the bottom of the unit and is slightly larger with an average value of 12%, than in the upper part of the unit (< 10%). The NCIR increases slightly from the bottom of the unit (85%) to the top (92%). The grain size distribution in unit 1 seem to be quite homogeneous, although a higher concentration of particles with sizes of approximately 20 μm occur in the upper part of the unit. The Ti intensities are higher and vary more than in the other units. The Si_{Ox} values are very low and increase slightly in parallel with SiO_2 content. Both Al_{Ox} and the Al_{Ox}/Al_{Tot} ratio are higher in the bottom part of the unit and decrease from the bottom to the top. Fe_{Ox} and the Fe_{Ox}/Fe_{Tot} ratio increase from the bottom part of Unit 1 to a depth of 0.7 m and then decrease to the top. The Fe_2O_3/Al_2O_3 ratio displays important variations in the bottom part of unit 1 and is more nearly constant in the upper part. The SiO_2/TiO_2 ratio varies rapidly and increases in the upper 0.15 m.

3.4. Age-depth model

A total of 9 radiocarbon analyses were performed on selected macrofossils collected from the 2.4-m length of the sediment core (Table 1). Three radiocarbon dates were excluded from the age–depth model (bold in Table 1). Samples Poz-73372 and Poz-73373, which were dated to 8000 and 4615 BP, respectively, are clearly as too old: their macrofossils had probably been reworked, and the corresponding ages have been removed from the age–depth modeling. The third excluded sample (Poz-73374) was associated with a change in the sedimentation rate that also corresponds to the change of sedimentary units (units 2 and 1). Thus, for the age–depth model constructed using the Clam routine, this date was excluded to better reflect the change between these sedimentary units (Fig. 5-a). The model was computed with a smooth relation until SacA-32339 and interpolated below.

The measurement of short-lived radionuclides allowed the dating of the uppermost 7 cm of the core (Appleby and Oldfield, 1978). A logarithmic plot of $^{210}Pb_{ex}$ activity (Fig. 5-b) shows a downward decrease from the surface. According to the 'constant flux, constant

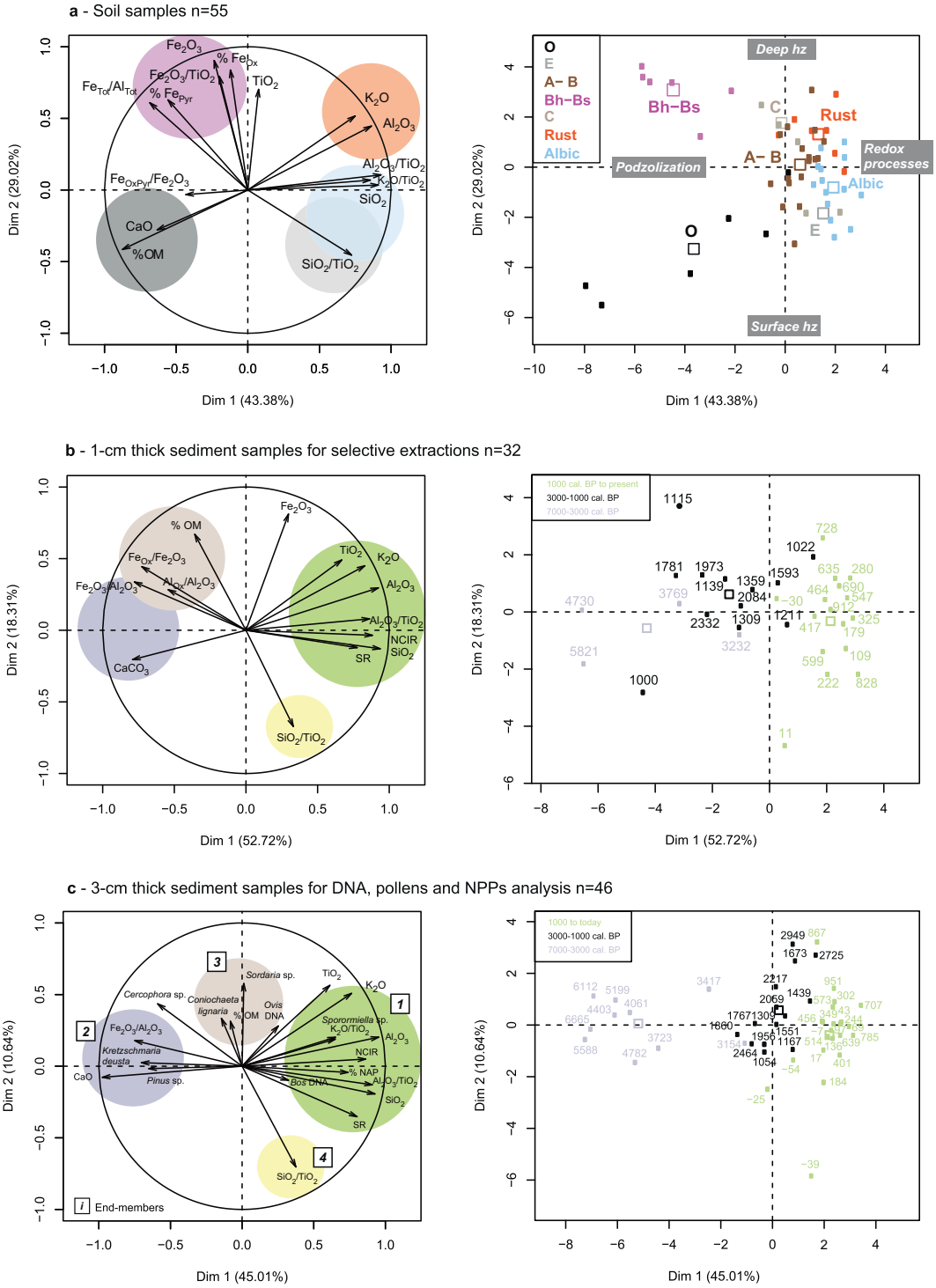


Fig. 3. Correlation circles and biplots of Dimensions 1 and 2 of the principal component analysis applied to data from the soil samples including paleosols (A), 1-cm-thick sediment samples (B) and 3-cm-thick sediment samples (C). Rust and Albic samples are horizons of Stagnosols, Bh-Bs are spodic horizons of Podzols, O are organic-rich horizons, A–B are poorly differentiated brown horizons and C samples are horizons of weathered rock. Numbers in the biplots shown in (B) and (C) represent the ages of the sediment samples. SR = sedimentation rate, hz = horizons.

sedimentation rate' (CFCS) model (Goldberg, 1963; Krishnaswamy et al., 1971), the levels of $^{210}Pb_{ex}$ indicate a mean accumulation rate of 0.7 mm·yr^{-1}. The ^{137}Cs activity profile (Fig. 5-b) has a peak at a depth of 4.05 ± 0.25 cm and reaches a maximum at 2.05 ± 0.25 cm, interpreted respectively as the beginning of atmospheric production of ^{137}Cs, in 1955 (due to nuclear tests) and the Chernobyl accident in 1986 (Robbins and Edgington, 1975). The expected high ^{137}Cs activity associated with the maximum intensity of nuclear weapons testing in the Northern Hemisphere in AD 1963 may lie between these two depths; however, it is not clearly identifiable likely due to both the low resolution of the analyses and the large amount of fallout from Chernobyl. The good agreement between the ages of the artificial radionuclide peaks inferred from the 210Pbex-CFCS model (1956 and 1984) and the actual ages of artificial radionuclide peaks (1955 and 1986) suggests that these dates provide an accurate and continuous age–depth relationship within the upper 7 cm of the sediment sequence. These data

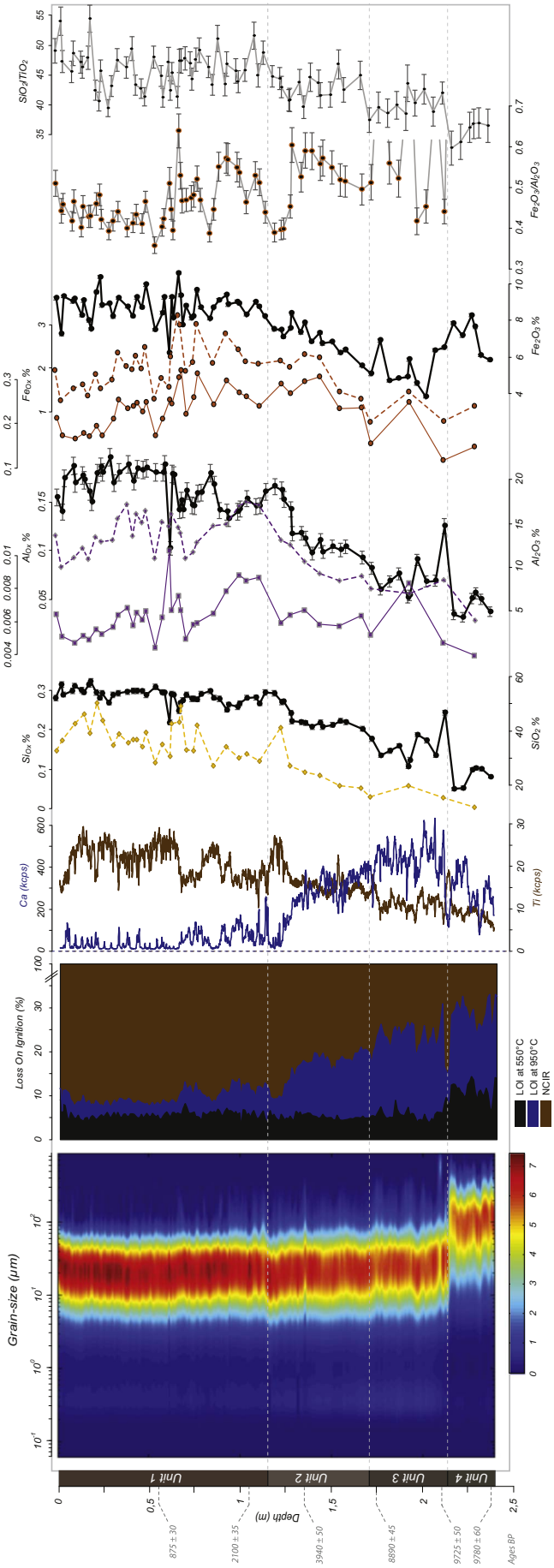


Fig. 4. Comparison of the grain-size of the sediment of Lake Verney according to depth with loss on ignition at 550 °C and 950 °C, its residue (NCIR), the XRF core scanner intensities of Ti and Ca and the XRF geochemistry of major elements (SiO_2, Al_2O_3 and Fe_2O_3, black line) with the results of ammonium oxalate extraction of these elements (Si_{Ox}, Al_{Ox} and Fe_{Ox}, dotted line) and their corresponding ratios (Al_{Ox}/Al_{Tot} and Fe_{Ox}/Fe_{Tot}) and the Fe_2O_3/Al_2O_3 and SiO_2/TiO_2 ratios. The color scale of the grain-size contour plot refers to the abundance of each grain-size class expressed as a percentage, with the least abundant shown in blue and the most abundant shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

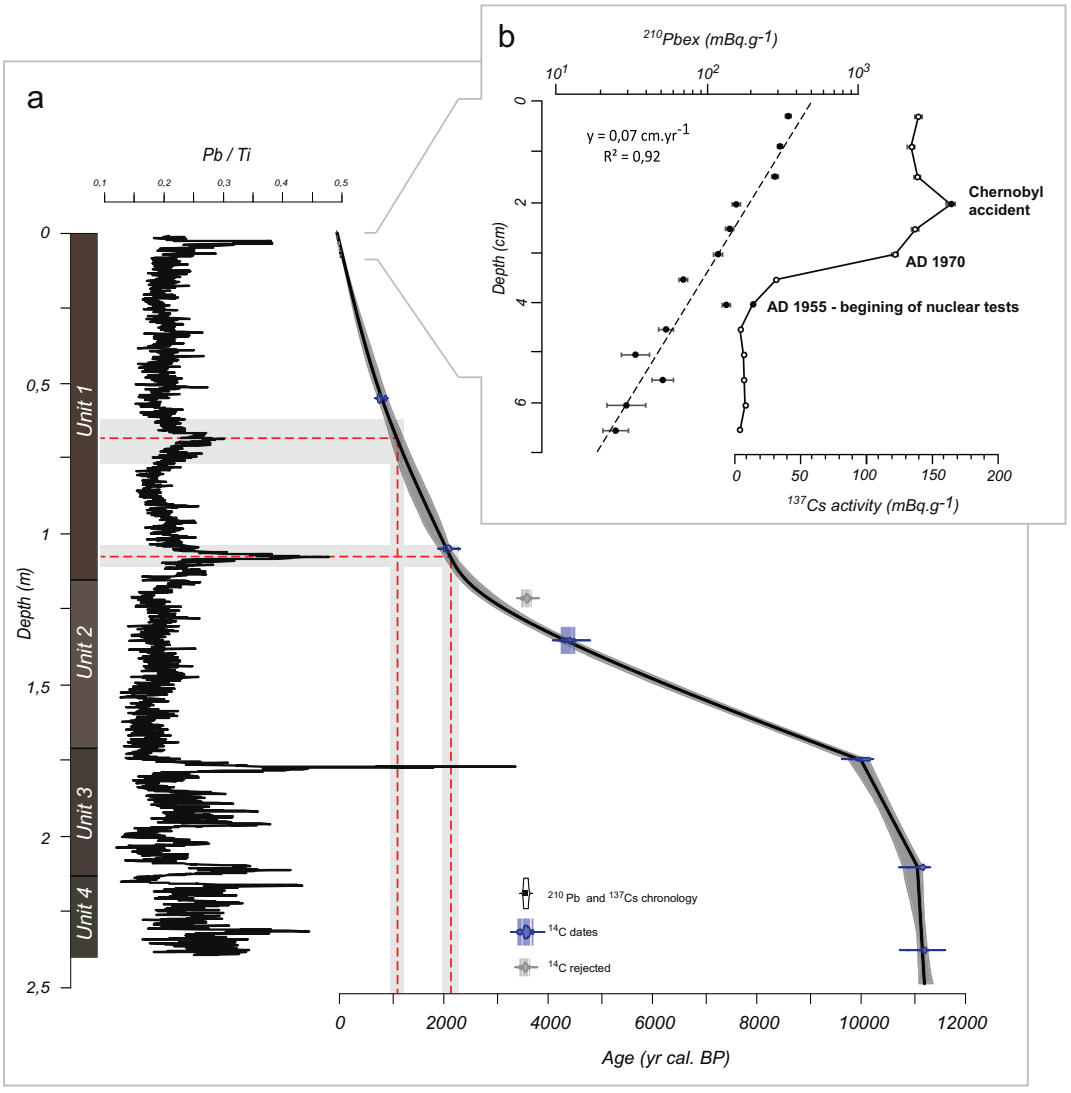


Fig. 5. Age-depth model of the Lake Verney sediment sequence based on ^{14}C (A) and $^{210}Pb/^{137}Cs$ ages (B). Uncertainties are included in the dots representing ^{137}Cs activity measurements. The Pb/Ti ratios enable comparison of the model with the Pb emissions during modern, Medieval and Roman periods.

were incorporated into Clam to generate the age–depth model of the whole core alongside the ^{14}C age estimates (Fig. 5-a).

The age–depth model shows three major phases with differing sedimentation rates within the 2.4 m of the sediment sequence (Fig. 5-a). Units 4 and 3 span the period from 11,200 to 9200 cal. yrs BP and have a mean sedimentation rate of 0.7 mm·yr^{-1} (Fig. 6-a). Unit 2 covers the period between 9200 and 2400 cal. yrs BP and has a mean sedimentation rate of 0.07 mm·yr^{-1}. The sedimentation rate of unit 1 increases from 0.2 to 0.7 mm·yr^{-1} over the last 2400 cal. yrs BP up to the present (Fig. 6-a).

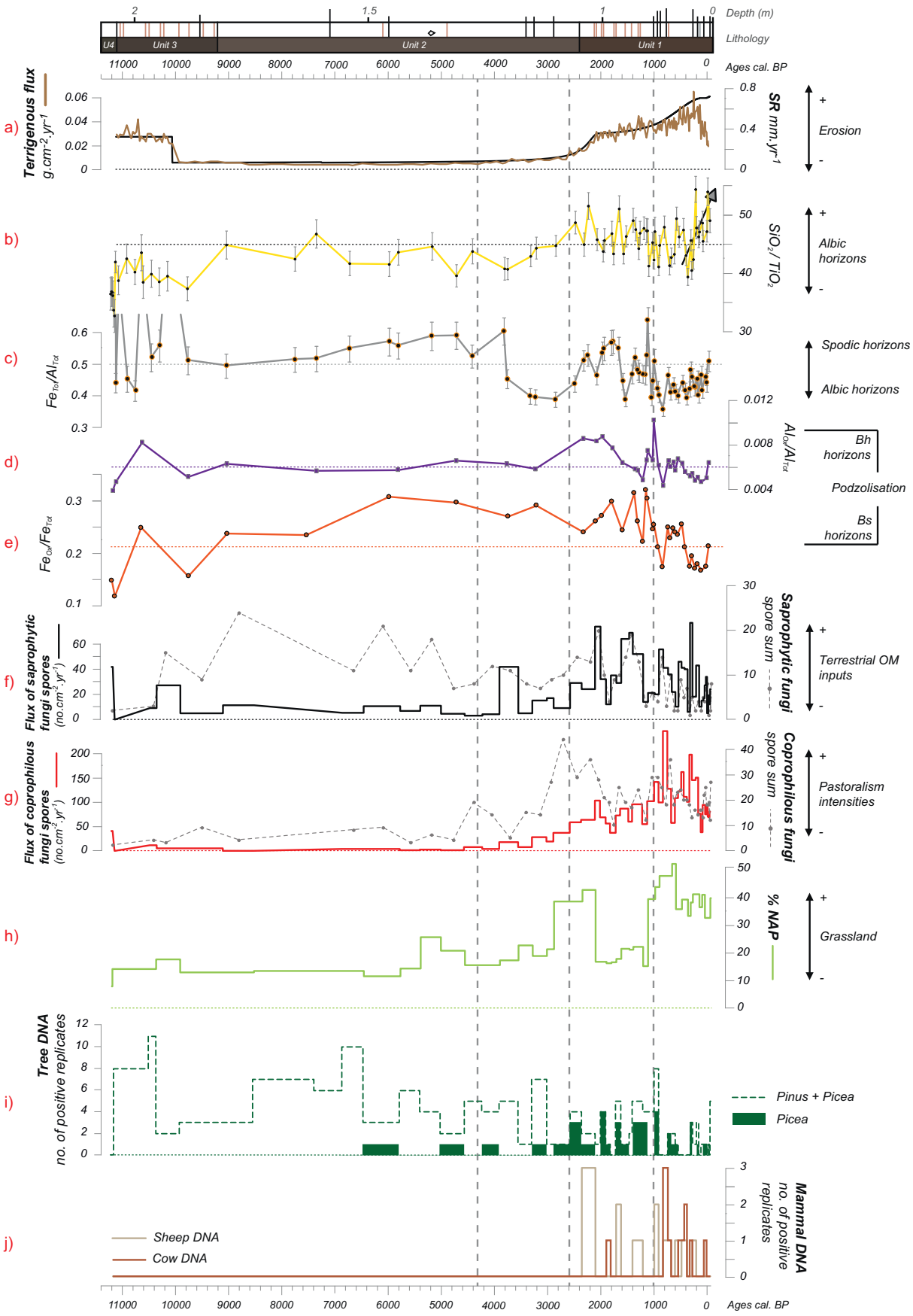
The variations in the values of the Pb/Ti ratio measured using the XRF core scanner show four periods of increases in Pb intensity (Fig. 5-a). At the bottom of the core, before 10,000 cal. yrs BP, Pb is considered to be linked to the organic matter within the sediment. After 10,000 cal. yrs BP, the three other increases are consistent with the Roman period, the Medieval period and modern emissions of leaded gasoline, respectively, and these peaks support our age-depth model (Arnaud et al., 2004; Elbaz-Poulichet et al., 2011).

3.5. Quantification of non-arboreal pollen (NAP) and NPPs

On average, NAP represents 16% of the total pollen from 11,400 to 3000 cal. yrs BP (Fig. 6-h). It increases from 3000 to 2000 cal. yrs BP to a maximum of 42%. NAP is once again at an average of 16% from 2000 to 1100 cal. yrs BP and increases again (up to 50%) after 1100 cal. yrs BP.

Among NPPs, we focus only on fungal ascospores produced in the catchment that grow on plant remains, within the soil litter, and in fecal material. *Cercophora* sp. (HdV-112), *Coniochaeta ligniaria* (HdV-172) and *Kretzschmaria deusta* (HdV-44) are assumed to be saprophytic fungal ascospores associated with the degradation of the litter associated with deciduous trees (Doyen and Etienne, 2017;), whereas *Sordaria* sp. (HdV-55B), *Sporormiella* sp. (HdV-113) and *Podospora* sp. (HdV-368) are saprophytic/coprophilous or strictly coprophilous fungal ascospores (Etienne and Jouffroy-Bapicot, 2014; Doyen and Etienne, 2017; Etienne et al., 2013). NPP types were summed relative to their saprophytic or fecal ecological indicator value and expressed in abundances (no·cm^{-3}) and accumulation rates (no·cm^{-2}·yr^{-1}) (Fig. 6-f-g).

The abundances of saprophytic spores increase from 11,400 cal. yrs BP to a maximum at approximately 9000 cal. yrs BP. They decrease progressively until 3200 cal. yrs BP. On the other hand, accumulation rates remain low during the early and mid-Holocene, except between 3500 and 3900 cal. yrs BP (Fig. 6-f). The accumulation rates of saprophytic NPPs and their abundances follow the same trend during the last 3000 years, with four major increases occurring from 2600 to 2000 cal. yrs BP, 1600 to 1300 cal. yrs BP, 900 to 700 cal. yrs BP and



(caption on next page)

Fig. 6. Evolution of pedo-landscapes and agro-pastoral activities in the Lake Verney catchment during the Holocene as inferred from a) sedimentation rates (SRs) and terrigenous fluxes (including 950 °C LOI and NCIR of the sediment); b) SiO_2/TiO_2; c) Fe_{Tot}/Al_{Tot}; d) Al_{Ox}/Al_{Tot} and e) Fe_{Ox}/Fe_{Tot} ratios to characterized the different sources of the soil horizons. The sum of saprophytic NPPs, including *Coniochaeta ligniaria*, *Cercophora* sp. and *Kretzschmaria* spores refers to the origin of the OM (f). Dung residue spores (g) represents the sum of *Sordaria* sp., *Sporormiella* sp. and *Podospora* sp. NAP is the sum of non-arboreal pollens (h). Tree DNA includes *Pinus* sp. and *Picea* sp. positive replicates (i), and mammal DNA (j) includes cow or sheep DNA positive replicates.

500 to 300 cal. yrs BP.

The abundances and accumulation rates of coprophilous ascospores both remain low until 4400 cal. yrs BP (< 10 spores and < 20 spore·cm^{-2}·yr^{-1}, respectively (Fig. 6-g)). The abundances of these spores reach 20 spores between 4400 and 4000 cal. yrs BP and a maximum value between 3000 and 2000 cal. yrs BP (45 spores). They remain high during the last 2000 years (10 to 35 spores), with small decreases occurring between 1800 and 1200 cal. yrs BP and 300 and 100 cal. BP. The accumulation rates increase progressively from 4000 to 800 cal. yrs BP, up to 250 spore·cm^{-2}·yr^{-1}, and remain high (100 to 200 spore·cm^{-2}·yr^{-1}), with short-term decreases occurring at approximately 700 cal. yrs BP and during the last 150 years (Fig. 6-g).

3.6. Sedimentary DNA

The DNA of five mammalian taxa was found in the sediments of Lake Verney, specifically *Bos* sp. (cattle), *Equus* sp. (horse), *Felis* sp. (cat), *Ovis* sp. (sheep) and *Sus* sp. (pig). *Equus*, *Felis* and *Sus* were only detected in one or two samples within the sequence. Because these animals are uncommon in alpine pasture contexts, we focus only on DNA of *Bos* sp. and *Ovis* sp. (Fig. 6-j). No DNA from either *Ovis* or *Bos* was detected in sediments older than 2360 cal. yrs BP. DNA of *Ovis* sp. was found in 10 samples from 2360 to 200 cal. yrs BP. In samples with ages ranging from 2360 to 900 cal. yrs BP, sheep DNA was generally detected in multiple replicate PCR analyses per sample, whereas it was only detected in one out of 12 PCR replicates after 900 cal. yrs BP. DNA of *Bos* sp. was only detected in one PCR replicate between 2360 and 900 cal. yrs BP. After 900 cal. yrs BP, *Bos* DNA was found in 8 samples, in 1 to 3 positive PCR replicates per samples.

There are currently no trees in the catchment of Lake Verney. Among the DNA of trees found in the sediments, *Pinus* is a pioneer species in the Alps whereas *Picea* became established in the region at *ca.* 3000 yrs BP (David, 1993; de Beaulieu et al., 1993). *Pinus* DNA was found in at least one replicate in all samples with ages between 11,200 and 2900 cal. yrs BP, with an average of 5 positive replicates per sample (Fig. 6-i). During the second part of the Holocene, the number of positive replicates decreased. It was then found in between 0 and 2 replicates since 2900 cal. yrs BP, except between 1000 and 900 cal. yrs BP and in the most recent sample (4 positive replicates). *Picea* DNA appears punctually at *ca.* 6000, 5000 and 4000 cal. yrs BP with just one positive replicate (Fig. 6-i). After 3000 cal. yrs BP, it was detected more often, with up to 4 positive replicates per sample, especially between 2600 and 1000 cal. yrs BP.

3.7. Statistical endmembers

In order to avoid the influence of the bottom of the sediment sequence, which shows substantial variations in composition and sedimentation rates, PCA analyses were only performed on sediments with ages of 7000 years or less.

The first two components of the PCA performed using the data set of sediment selective extractions (n = 32) represent 71% of the variability (Fig. 3-b). The arrangement of variables according to both dimensions is similar to the PCA results from the soil samples. However, OM is more closely related to variations in iron, and it is positively correlated with the first component and negatively correlated with the second one, and the Fe_2O_3 and TiO_2 contents appear to be drawn toward the axis of the first component and its terrigenous endmembers. The sedimentation rate (SR) and NCIR are also related to this group, supported by the first component. SiO_2/TiO_2 is always positively correlated with the first component and negatively with the second one, while the calcium carbonate content is negatively correlated with both components. Finally, the values of the Al_{Ox}/Al_{Tot} ratio in the sediment seem to follow the same changes in the iron ratios and the OM contents.

The first two dimensions of the PCA performed using the data set including NAP, *Pinus* DNA and spore abundances of *Kretzschmaria deusta*, *Cercophora*-type, *Coniochaeta ligniaria*, *Sordaria* sp. and *Sporormiella* sp. represent 56% of the total variability (Fig. 3-c), and the resultst of this PCA appear to reproduce the same scheme as the PCA of the data set including selective extraction of Fe and Al (Fig. 3-b). Four endmembers can be identified. NAP, which is positively correlated with the first component, is strongly linked to the elements of terrigenous origin and ratios involving those elements. In contrast, CaO content, *Pinus* DNA, the spores of *Kretzschmaria deusta*, Fe_2O_3 and Al_2O_3 form a second group and are negatively correlated with the first group. A third endmembers that is associated with OM content and *Sordaria* and *Coniochaeta ligniaria* spores show a weak positive correlation with the second component of the PCA. Spores of the *Cercophora*-type are located between these two last endmembers and are positively correlated with the first component and negatively with the second one. As *Ovis* and *Bos* DNA occur sporadically in the sediment, their representativeness is poor in the PCA. *Ovis* DNA seems to be associated with both OM and terrigenous inputs. On the other hand, *Bos* DNA is especially closely related to the terrigenous endmembers and the sedimentation rate and by the SiO_2/TiO_2 ratio. These components make up the fourth endmember, which is positively correlated with the first component and negatively correlated with the second one. The spores of *Sporormiell*a sp. are situated between the directions of *Ovis* and *Bos* DNA, which is consistent with their relationship to pastoralism (Etienne and Jouffroy-Bapicot, 2014; Doyen and Etienne, 2017).

The biplot mapping of the results of both sediment PCAs according to the ages of the sediment samples shows a gradation along the first component, from the oldest sediments (7000–3000 cal. yrs BP), which are negatively correlated with the first axis, to the more recent sediment (1000 cal. yrs BP to the present), which are positively correlated with the first axis (Fig. 3-b-c). The sediments of the period between 3000 and 1000 cal. yrs BP are situated in the middle of the first axis and seem to be more closely associated with the second component.

4. Discussion

4.1. Sediment origins

Sediment inputs in the lake represent both changes in the pedo-landscape around Lake Verney and the diversity of that pedo-landscape, just as the current soils do, and areas of preferential erosion. These sediments continuously record a mix of different sources, depths and degrees of weathering, including, in a global way, the expected dominant processes. They also record the main eroded areas that are the steepest slopes, mainly upstream, surrounding the lake and areas where erosion is induced by human activities. The steep slopes concern only Regosols or Leptosols that are expected to provide always the same constant contribution through time, downstream and to the sediment record. The study of erosion before the development of human activities, *i.e.*, on long time-scale, allows us to dissociate this erosion source from the expected human induced erosion, assuming that other more advanced soil types are few eroded in the absence of human activity. Our goal is to differentiate the main sources of material, referring to the main pedogenetic processes that have occurred in the catchment. Both carbonate schist in the west and black shales in the east can provide

calcium (through dissolution) and calcium carbonates (through physical erosion) in the lake. Calcium carbonates recorded in the sediment may come from both terrigenous and authigenic inputs, but both inputs can reflect early conditions of soil development (decarbonatation), with small inputs of weathered material or the loss of carbonate material from young soils at the beginning of pedogenesis and associated with the rejuvenation of soils, *e.g.* on steep slopes (Bajard et al., 2017). The OM contents recorded in the sediments are constant over the last 10,000 yrs, whereas the sediment flux has increased. This result suggests that the OM was derived from the catchment and was not produced authigenically in the lake. The terrestrial origin of the OM is supported by the record of saprophytic fungi spores, which show comparable abundances before and after the increase in erosion (Fig. 6-f).

4.2. Record of soil footprints in the sediments

Identification of the soils that fed the lake sediments is made possible by the characterization of current soils on both hillsides of the pass and by the high differentiation of horizons between the soil types identified and within the same soil. Both the soils of the lake catchment and the soils close to the lake catchment were compared to the sediment record in order to catch all pedogenetic processes that could no longer operate in the lake catchment. Soils and especially paleosols were also considered to validate our interpretation from the sediment sequence. The development of undifferentiated and thin Regosols on limestone, whereas the other soils are acid, suggests that the decarbonatation from rock to soil during weathering is relatively fast. Only the coarse fragments in the matrix effervesce to HCl solution, and these elements can provide terrigenous carbonates to the lake. This quick acidification also allowed the development of Podzols on material that had previously been leached of carbonate (*e.g.*, PSB3 and PSB4) over relative short timescales (*e.g.*, Bormann et al., 1995).

Both Podzols and Stagnosols are very acid soils (pH 4 to 5, *e.g.*, PSB3 and VERS2); thus, they can develop from acid Cambisols. Both are characterized by different horizons, albic and rust in particular.

4.2.1. Identified albic horizons

The albic horizons of both the Podzols and Stagnosols are characterized by a light color ("bleached" or "whitened") and are enriched in silica (high ratios of SiO_2/TiO_2), reflecting the eluvial character of these horizons (Figs. 3-a and 7). In consequence, they are difficult to distinguish, except for the E horizons of Podzols. These horizons occur between OM-rich horizons; thus, they are more often associated with organic matter than albic horizons, which are conversely more closely associated with strictly terrigenous materials, such as SiO_2, K_2O and Al_2O_3 (Figs. 3-a and 7).

4.2.2. Differentiation of the rust horizons of Stagnosols from the spodic horizons of Podzols

The rust horizons of Stagnosols and the spodic horizons of Podzols are well differentiated by different variables with regard to the PCA performed on the soil samples (Fig. 3-a). The spodic horizons are characterized by high values of Fe_{Ox}, Fe_{Pyr}, and Fe_{Tot}/Al_{Tot}, whereas the rust horizons of Stagnosols are more strongly characterized by K_2O and Al_2O_3. Organo-metallic complexes of Al and Fe form in the sub-surface horizons of Podzols and accumulate in the Bh and Bs horizons (Fig. 7). Al-OM complexes can accumulate to a greater degree in Bh horizons under hydromorphic conditions (Anderson et al., 1982; Legros, 2007). Based on the studies of Mourier et al. (2008, 2010), these geochemical proxies are the characteristics that we will look for in the sediment, assuming their conservative behavior during mobilization, transport and deposition from the catchment to the lake and that they did not form in the lake (Fig. 7). This assumption is possible considering the small size of the lake catchment, and the smaller areas in the catchment concerned by soils and not only scree. Climatic and pH conditions are also in favor of low biological and chemical changes of soil-sediment properties within the lake (*i.e.*, low temperatures and higher pH in soils than in the lake).

Only extractions with ammonium oxalate solution were performed on sediment samples. Fe_{Ox}, Al_{Ox}, and Si_{Ox} represent the amount of both amorphous and complexed forms, whereas extractions with Na-pyrophosphate only extract the complexed forms. The Podzols are characterized by the presence of OM-complexes and the small differences between Fe_{Pyr} and Fe_{Ox} seen in PSB3 and PSB4. These results indicate that little or no amorphous material was present in the Podzols. The very low contents of Si_{Ox} in the sediment indicate that there are no amorphous compounds of Si, and thus we assumed that there are also no amorphous compounds of Fe and Al, as we did for the horizons of VERS1 and VERS2. Furthermore, the microbial activity of the lake could change Fe-OM complexes into amorphous compounds (Birkeland, 1999; Mourier et al., 2010). In consequence, the contents of Fe_{Ox} and Al_{Ox} are attributed to the OM-complexes of spodic horizons. Finally, as the sediment reflects a mixture of different sources, we normalized Fe_{Ox} and Al_{Ox} to the contents of total iron (% Fe_2O_3) and aluminum (% Al_2O_3), respectively, to trace the inputs from spodic horizons into the lake (Fig. 7). In the PCA performed using data from the selective extractions of sediments (Fig. 3-b), both the Fe_{Ox}/Fe_{Tot} and Al_{Ox}/Al_{Tot} ratios are strongly correlated with the Fe_{Tot}/Al_{Tot} that strongly characterizes the spodic horizons, thus supporting our interpretation.

4.3. Evolution of soils and agro-pastoral activities throughout the Holocene, as inferred and dated from sediment archives of Lake Verney

4.3.1. 11,000 to 4300 cal. BP: from decarbonatation of the catchment to acidification of soils

The composition and chronology of unit 4, which was deposited between 11,200 and 11,100 ± 100 cal. yrs BP suggests instantaneous deposition of this unit, similar in its carbonated nature to the parent material, at the beginning of the lake sedimentation (Figs. 1, 4 and 5-a). The sedimentation dated to 11,100 and 10,000 cal. yrs BP is rapid but lower than the sedimentation rates seen in unit 4 and displays a high variability in the OM contents and the values of the Fe_2O_3/Al_2O_3 ratio.

Between 10,000 and 4300 cal. yrs BP, erosion was lower, and *Pinus* DNA became abundant (Fig. 6-i). The high abundance of charcoals of *Pinus cembra* found in four soil profiles in the area of the pass confirms that the DNA recorded in Lake Verney come from this species (Talon, 2006). Five of them were found at 1940 m on the French side of the pass and were dated to between 4325 and 3910 cal. yrs BP (Talon, 2006), in agreement with the chronology of *Pinus* DNA in the sediment (Fig. 6-i). The abundance of saprophytic spores between 10,000 and 4300 cal. yrs BP, confirms the development of litter horizons under the forest (Fig. 6-f). At the same time, the carbonate contents decrease and the values of Fe_{Ox}/Fe_{Tot} ratio increase, suggesting that organic acids from the forest litter contributed to the dissolution of the soil carbonates and acidified the soils, permitting dissolution of Fe, which formed complexes with OM compounds and migrated through the soils, and could be laterally transported into the lake. The change in carbonate contents in the early to mid-Holocene sediment appear to have been controlled mainly by weathering of the catchment, favored by forest covers (Giguet-Covex et al., 2011; Dreibrodt and Wiethold, 2015; Bajard et al., 2017). The values of Al_{Ox}/Al_{Tot} ratio do not vary within the sediment deposited during this period, perhaps due to the lower mobility and lower contents of Al_{Ox} compared to Fe_{Ox}. In consequence, it is possible that the Al complexes migrated within the soils, as did the Fe complexes, forming spodic horizons but not reaching the lake. The development of spodic horizons is corroborated by the increase in the values of the Fe_{Tot}/Al_{Tot} ratio in the lake sediment between 10,000 and 4300 cal. yrs BP. Moreover, the time between the beginning of pedogenesis and 4300 cal. yrs BP seems long enough to allow the differentiation of such horizons (Bormann et al., 1995; Protz et al., 1984). Pedogenesis may have been favored by the climate warming that

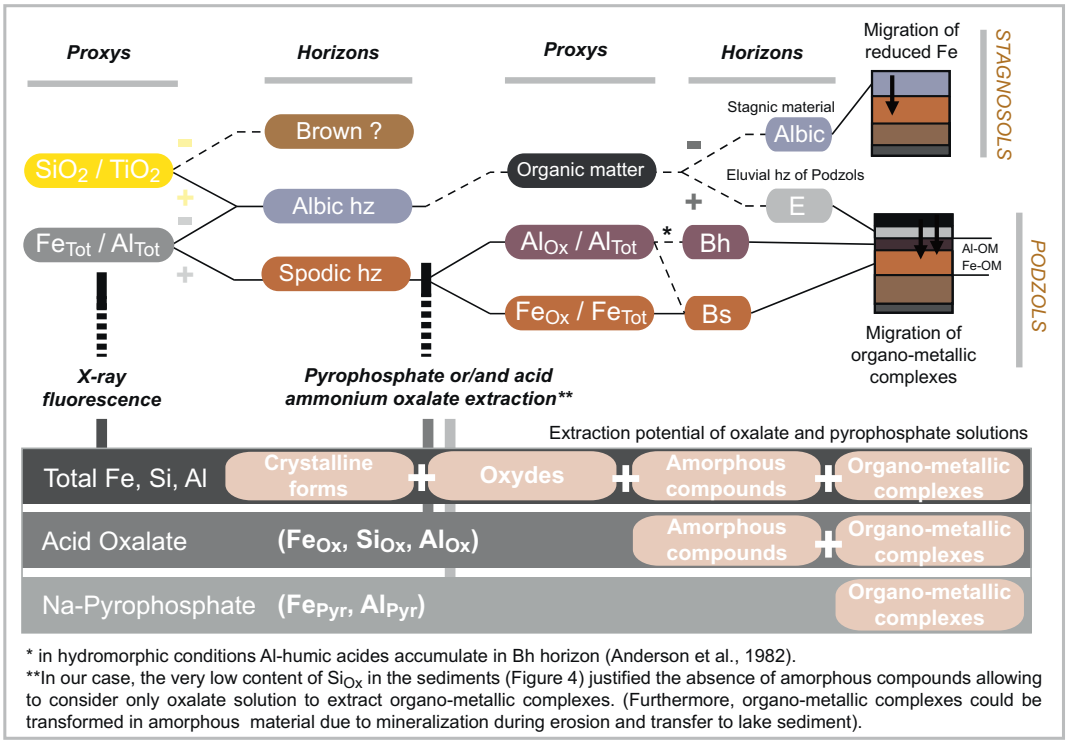


Fig. 7. Identification scheme of the horizon signatures of Podzols and Stagnosols in lake sediments from SiO_2/TiO_2, Fe_{Tot}/Al_{Tot}, Fe_{Ox}/Fe_{Tot} and Al_{Ox}/Al_{Tot} ratios. The Al and Fe-organo-metallic complexes that characterize the spodic horizons of Podzols were extracted with the acid ammonium oxalate solution. Hz = horizon, Ox = oxalate, Pyr = pyrophosphate.

occurred during this period, which is assumed to have reach its maximum between 6000 and 7000 cal. BP (Millet et al., 2008). Thus, following the loss of carbonate compounds, acidification led to the development of Podzols in the catchment of Lake Verney, where no Podzols are currently present. Neither DNA nor NPPs indicate human activities in the Lake Verney catchment before 4300 cal. yrs BP.

4.3.2. 4300 to 2600 cal. BP: first human occupation in the catchment: deforestation and litter erosion

From 4300 cal. yrs BP, erosion increased gradually from 0.004 to 0.008 $g{\cdot}cm^{-2}{\cdot}yr^{-1}$ and forest reduction and inputs of coprophilous fungi spores represent the first opening of the landscape to human activities in the catchment (Fig. 6-a-h-i). Fire episodes inferred from the anthracological study of Talon (2006) are consistent with forest clearing at this altitude and during this period. The migration of *Picea* into the area of the pass may have been favored by human activities (Miras et al., 2006).

The variability in *Kretzschmaria deusta* and *Coniochaeta ligniaria* was related to the *Pinus* DNA and the OM contents, respectively (Fig. 3-c) and suggest an erosion of litter and soil organic horizons between 4000 and 3600 cal. yrs BP (Fig. 6-f). Erosion of surface soil horizons could explain the substantial amount of *Pinus* DNA detected when clearing was initiated. On the other hand, the absence of mammal DNA in the sediment could reflect a low intensity of pastoral activities.

After 3800 cal. yrs BP, the input of albic material increased (Fig. 6-b). Following the erosion of the organic horizons, erosion of the eluvial horizons of the Podzols seems to be the most convincing hypothesis for this increase, but we cannot exclude erosion of the albic horizons of the Stagnosols. Indeed, the albic and rust horizons of the Stagnosols are more closely associated with TiO_2, Al_2O_3 and SiO_2 than the E horizons of the Podzols (Fig. 3-a), suggesting that the first record of stagnic horizons occurred between 2900 and 2600 cal. yrs BP (Figs. 4 and 6).

4.3.3. 2600 to 1000 cal. BP: erosion of spodic horizons

During the Second Iron Age and the Roman period, pastoral activity steadily increased, as reflected by the occurrence of substantial amounts of NAP, whereas the detection of tree DNA diminishes (Fig. 6-g-h-i). According to the mammal DNA, more sheep flocks than herds of cows were going up to pasture around the pass during this period (Fig. 6-j). In association with this intensification of pastoral activities, the erosion rate strongly increased between 2600 and 1900 cal. yrs BP, as both of the Al_{Ox}/Al_{Tot} and Fe_{Tot}/Al_{Tot} ratios, followed by the increase in Fe_{Ox}/Fe_{Tot} after 1900 cal. yrs BP, indicating the successive erosion of the Bh and Bs spodic horizons of the Podzols (Fig. 6-a-c-d-e). This result indicates that the erosion of the albic horizons during the previous period was more closely associated with eluvial horizons than the albic horizons of Stagnosols. The variability in the values of SiO_2/TiO_2 and Fe_{Ox}/Fe_{Tot} ratios between 1900 and 1000 cal. yrs BP seems to indicate mixing with material from albic horizons, especially between 1600 and 1400 cal. yrs BP when lower values of the Fe_{Tot}/Al_{Tot} ratio are noted. Except for SiO_2/TiO_2, the values of all the ratios (*i.e.*, Al_{Ox}/Al_{Tot}, Fe_{Ox}/Fe_{Tot} and Fe_{Tot}/Al_{Tot}) began to decline around 1900 cal. yrs BP, suggesting a change in pedogenetic processes, from podzolization to the development of stagnic properties. Erosion of the Bh and Bs horizons is supported by the occurrence of the two periods during which the largest terrigenous fluxes took place and by the terrestrial OM inputs (Fig. 6-a-f), which could reflect the organic character of the eroded horizons, and could have been transported with the Al and Fe-OM complexes (Wilmshurst and McGlone, 2005).

Between the Antiquity and the Middle Ages, during the so-called "Dark Ages", a significant decrease in the NAP associated with an increase in the tree DNA was recorded, indicating recolonization of the site by pine (likely *Pinus cembra*) and spruce (Fig. 6-h-i). In a few short centuries, this recolonization could have led to a recovery in podzolic processes as suggested by the slight increases in the spodic ratios (*i.e.*, A_{lOx}/Al_{Tot}, Fe_{Ox}/Fe_{Tot} and Fe_{Tot}/Al_{Tot}). This recovery of the forest is supported by both human and climate forcing. Hearths found around the Roman building were dated to between 1834 and 1580 cal. yrs BP (Rey et al., 2015). These dates are in accordance with the old coin inventory of Rémy et al. (2006), and suggest at least use of the pass at that

time. The destruction of the Roman building at approximately 1600 cal. yrs BP and the occurrence of a colder period recorded approximately 1700 cal. yrs BP (Millet et al., 2008) could be related to abandonment of the area and the beginning of this new, short-live pedogenetic phase.

4.3.4. 1000 cal. BP to present: development of stagnic properties

From 1000 to 100 cal. yrs BP, erosion rates continued to increase, from 0.04 to 0.06 $g \cdot cm^{-2} \cdot yr^{-1}$. Between 1000 and 400 cal. yrs BP, mammal DNA, NAP and the spores of coprophilous fungi spores reach their maxima (Fig. 6-h-i-j). They increase again between 400 cal. yrs BP and the present. During both of these periods, the occurrence of tree DNA remains sporadic, indicating that clearing was maximum, as it does today, without trees in the catchment. The most important change in pastoral activities is reflected by the most pronounced detection of *Bos* DNA that suggests an increase in cattle breeding after the Middle-Ages. Indications of podzolization become more and more sparse, and the development of stagnic properties has dominated pedogenesis since at least 400 cal. yrs BP.

4.4. Validation of soil evolution with current soils, paleosol and anthracological analyses

The current soils, the studied paleosols, plus the anthrocological study performed by Talon (2006), in the area of the Petit Saint-Bernard Pass as part of the Alpis Graia Interreg Project, provide elements that support the soil evolution scenario proposed using the sediment record from Lake Verney.

First, the current pedogenesis of the lake catchment, dominated by stagnic properties, is consistent with the interpretation of the latest sediment deposits. No evidence of Podzols was found in the lake sediment for the last 400 years cal. BP, in agreement with the current absence of Podzols in the catchment of Lake Verney. Podzols can develop at higher altitudes than those found on the French side of the pass (*e.g.*, PSB3) and on the black shales of the Houillere Briançonnais area and below the subalpine Rhododendron heaths, as reported by Mourier et al. (2008). The development of true Podzols is also possible on calcareous schists (obviously after decarbonatation), and in wet climates, such as those found in the Northern Alps (Legros and Martini, 1992). Podzols developed on Liassic calcareous schists have been described from the Balme Pass (2200 m), located 40 km to the north of the Petit Saint-Bernard Pass (Dambrine, 1985). Thus, the current absence of Podzols around Lake Verney cannot result from the altitude of the lake, which is above 2088 m, nor from the geological materials underlying the catchment.

Clear evidence of spodic horizons were identified in the sediment record, especially between the mid-Holocene period and the Middle Ages. Paleopodzols like those seen in PSB4 have been found at lower altitudes (2000 m) on the French side of the pass, indicating that old, but undated Podzols had developed and were subsequently covered by erosion induced deposits. Erosion increased beginning 4300 cal. yrs BP; thus, it is possible that Podzols were present before 4300 cal. yrs BP. Moreover, the deepest charcoals found within a Podzol (beyond the organic-mineral horizon) at 1940 m of elevation were dated to 1210 and 4160 cal. yrs BP (Talon, 2006). The other 8 dates obtained from this profile ranged from 2105 to 4325 cal. yrs BP, except for one sample that yielded an age of 330 cal. yrs BP and was obtained from the organic surface horizons. These charcoals indicate that fire episodes occurred during this interval, and that Podzols were present in the area of the pass before the Roman period and before the strong increase in erosion.

Among the successive albic and rust horizons of both paleosols PSB1 and PSB2, the deepest rust horizons display high contents of Fe_{Ox}, and slightly more Fe_{Pyr} was noted in PSB1 (at depth of approximately 30 and 40 cm). These characteristics are reminiscent of the spodic character of the Podzols, even if the values are 2 to 3 times less important than for Podzols (Fig. 2). The other albic and rust patterns do not display these characteristics. Our hypothesis is that these deep, Fe_{Ox}-enriched horizons represent old degraded spodic horizons. Removal of the surface horizons and their exposure could allow the mineralization and degradation of organo-metallic complexes and the formation of amorphous compounds, *e.g.*, ferrihydrite, suggested by the high ammonium oxalate extractable Fe fraction (Birkeland, 1999; Stucki et al., 1988). OM-complexed forms can oxidize due to drainage and form ferrihydrite. This process is characteristic of acid hydromorphic soils (Jeanroy, 1983). Hydromorphic processes may have been present in the Podzols but had only a weak effect while podzolization occurred.

Furthermore, if we consider the thicknesses of the Podzols currently found in the area of the Petit Saint-Bernard Pass, to the Bs horizons, *i.e.*, approximately 30 cm (PSB3, Fig. 2-a), and the period during which the Podzols were eroded, approximately 3000 yrs (between 4000 and 1000 cal. BP), the potential erosion rate is 0.1 $mm \cdot yr^{-1}$ on average. This order of magnitude falls within the range of erosion rates triggered by agricultural management over the last two millennia (Enters et al., 2008; Massa et al., 2012; Bajard et al., 2017). It is also in the commonly admitted range of soil formation rates (Montgomery, 2007; Larsen et al., 2014) and raises the question of the sustainability of the pasture, which appears to be in equilibrium with the activities taking place in the catchment.

Charcoals were only found in PSB1 and were dated. Six ^{14}C dates provide information on the different levels and hearths found within the soil (Fig. 2). The deepest horizons were dated to 4725 and 4705 cal. yrs BP. Just below, the deepest hearth was dated to 4405 cal. yrs BP. The second larger hearth was dated to 2915 cal. yrs BP. The two other rust levels above were dated to 2600 and 2545 cal. yrs BP. Both hearths were considered to represent sporadically occupied campsites. Nevertheless, they allowed determination of the ages of the rust and albic patterns at the same burial levels. The evidence of spodic horizons is thus older than 2600 cal. yrs BP. In the sediments, erosion of the spodic Bh horizons, followed by the Bs horizons was recorded after 2600 cal. yrs BP, which is consistent with the absence of these horizons in the soils after 2600 cal. BP. Dates obtained from potential old spodic levels in PSB1 that range from approximately 3000 and 2600 cal. yrs BP are also consistent with dates obtained from a Podzol by Talon (2006) at an elevation of 1940 m. Moreover, two other profiles at 2000 m on the Italian side of the pass (*La Thuile/Plan Veyle Verney 3–2000 m and La Thuile/Plan Veyle Verney 2–2004 m*) also contain hearths and albic levels that are contemporary with the recorded human activities (Rey et al., 2014). Thus, the stagnic horizons seem to be more recent than the spodic horizons, suggesting that hydromorphic processes began to dominate the Podzolization. A Roman tile found in the albic horizon of a current Stagnosol (*Col du Petit Saint-Bernard – Sondage 14*) confirms this trend (Rey and Moulin, 2006).

4.5. Pedo-landscape evolution around the Petit Saint-Bernard Pass and implication for the evolution of the ecosystem cycle

Following glacier retreat and lake formation, Cambisols are expected to have quickly developed and acidified in the area of the pass, helped by the growth of vegetation, particularly by *Pinus cembra* (Fig. 8-a-b-c). On the steeper slopes, colluvial movement of soils may have regularly rejuvenated these Cambisols, as seen in VERS3 (Figs. 2 and 8-e). Furthermore, within the upper part of the catchment, pedogenesis was likely slower, due to the sparser vegetation, longer periods of snow cover, steeper slopes and rock landslides that are found there. Thus, only weakly differentiated profiles are found in the western part of the catchment of Lake Verney, as reported by Talon (2006). As a consequence, Regosols and Cambisols were present in the landscape throughout the Holocene, though they may have been more strongly represented in the landscape before podzolization (Fig. 8-a-b-c).

The process of Podzolization seems to have progressed after 10,000 cal. yrs BP and likely reached its maximum between 6000 and

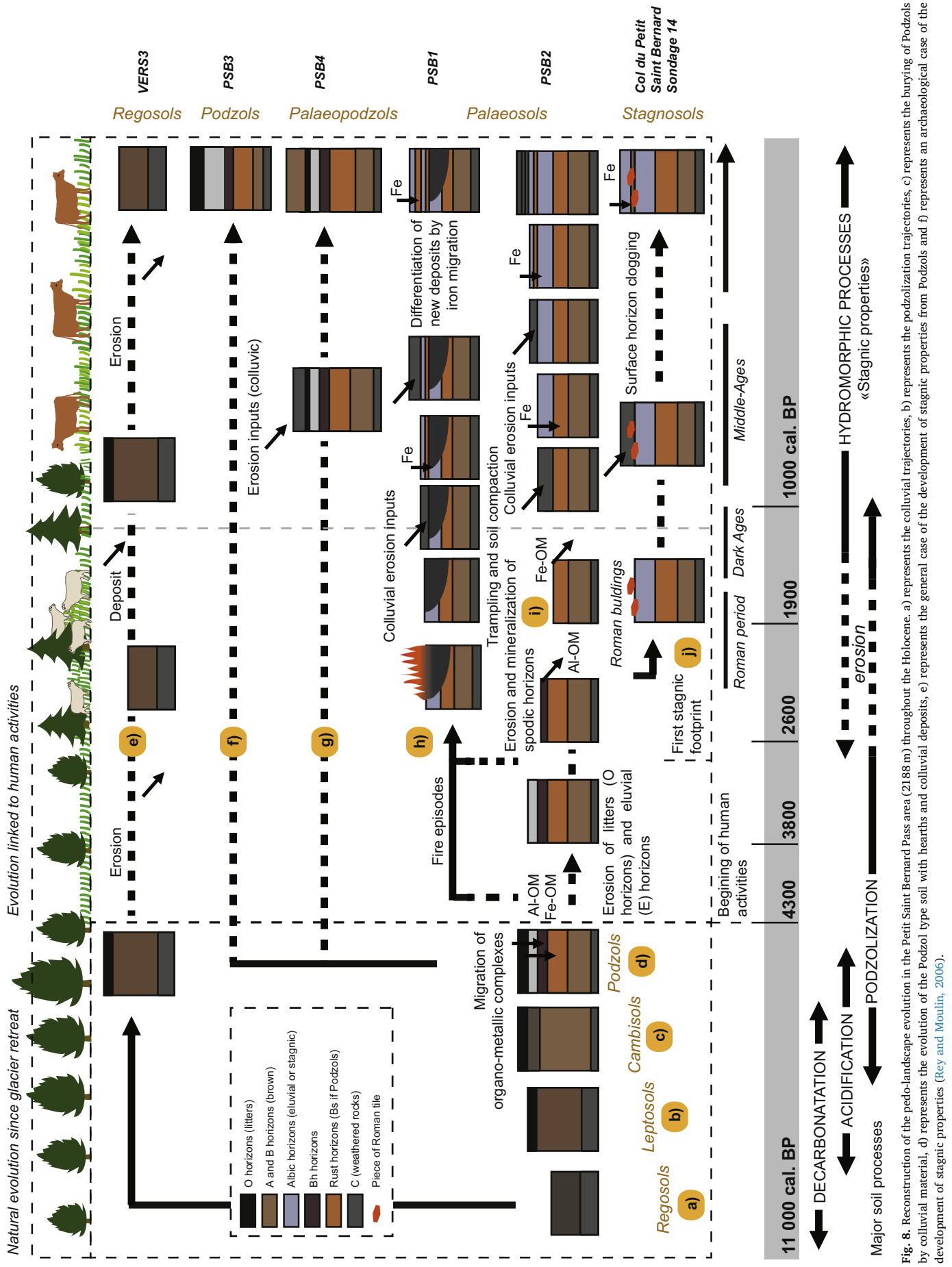


Fig. 8. Reconstruction of the pedo-landscape evolution in the Petit Saint Bernard Pass area (2188 m) throughout the Holocene. a) represents the colluvial trajectories, b) represents the podzolization trajectories, c) represents the burying of Podzols by colluvial material, d) represents the evolution of the Podzol type soil with hearths and colluvial deposits, e) represents the general case of the development of stagnic properties from Podzols and f) represents an archaeological case of the development of stagnic properties (Rey and Moulin, 2006).

4000 cal. yrs BP (Figs. 6 and 8-d). Following the beginning of human activities at approximately 4300 cal. yrs BP, changes in vegetation cover that were associated with an increase in erosion disrupted the soils that were present in the area of the pass. First, between 4300 and 2600 cal. yrs BP, soil litter and organic surface horizons were eroded, followed by the eluvial horizons of Podzols and parts of the spodic horizons between 2600 and 1000 cal. yrs BP (Fig. 8-h-i). A current analog was described by Pallmann (1947), who reported that, following thinning of a *Pinus cembra* forest, the settlement of grass and pasture led to a rapid degradation of acid humus horizons, transforming Podzols into poorly developed Podzols, grazing triggering a regression in soil evolution, and mixing the previously well-defined horizons.

As erosion increased, the spodic horizons came to the surface and mineralized, forming amorphous compounds. These compounds could be mobilized and transported or they could be covered with new materials (spodic or other) from upper parts of the slope (Fig. 8-h-i-j). In some cases, as PSB4, erosion deposited material that buried the former soils, consequently forming paleosols (Figs. 2 and 8-g) In the upper part of PSB2, regular inputs of coarse elements, slightly weathered, associated with higher pH of 8 would have led to the prevention of Fe migration and the differentiation of rust and albic horizons (Figs. 2 and 8-i). Regular deposits are much easier to mobilize than the fine material generated from calcareous schists which is prone to solifluction, generating humps and hollows (Legros and Martini, 1992).

Archaeological evidence, such as the hearths noted in PSB1 (Fig. 8-h) or the piece of Roman tile found in *Séez/Col du Petit Saint-Bernard – Sondage 14* (Fig. 8-j) are sometimes found between two levels of deposits. A new deposit is then emplaced producing new material for pedogenesis that dissociates between rust and albic levels due to the presence of vegetation. Indeed, change in litter, and new hydrological conditions in the soils of the catchment can prevent podzolization. Modification of hydrological conditions associated with pedogenetic change occurred due to the change in vegetation from native forest to land clearing (Willis et al., 1997; Mourier et al., 2010) but also due to the long use of the pasture over millennia and the possible development of irrigation in the area of the pass (Rey, 2011). Indeed, hydrological changes such as drainage can led to rapid changes in soil properties (Montagne et al., 2016). Several authors have described the degradation of Podzols, but mainly in term of evolution into Cambisols (Dambrine, 1985; Willis et al., 1997; Mourier et al., 2008; Poulenard et al., 2015). For example, Dambrine (1985) described the degradation of Podzols into Cambisols ǂ in response of the spread of agriculture into altitude pasturelands. The climate-specific conditions of the Petit Saint-Bernard Pass and the possible initially hydromorphic properties of the Podzols could have prevented from their development into Cambisols.

In the Middle Ages, the changes in the composition of herds, from primarily sheep to cows and/or the intensification of pastoralism appear to have been responsible for both the intensification of erosion and the acceleration of hydromorphic processes. Cattle can compact soils and cause increased runoff (Trimble and Mendel, 1995), changing the susceptibility of soils to water erosion (Giguet-Covex et al., 2011; Brisset et al., 2017). However, the depletion of soils in organic matter, due to the degradation of Podzols and the development of Stagnosols can also be responsible for degradation of the soil structure (Oldeman, 1994), and this process can accentuate erosion without any change in pasture intensity as a possible feedback process (Yaalon and Yaron, 1966).

To established a general model of soil evolution trajectories, we replace the soil evolution of the Petit Saint Bernard pass area in the ecosystem cycle evolution described by Holling (1986) and adapted for soil cycle functioning by Poulenard et al. (2015) (Fig. 9). The cycle is made up of 2 axes that correspond to *Capital versus Connectedness and organization* which were adapted to *Stock* and *Fluxes* for the soil system (Poulenard et al., 2015). We choose to address organic carbon here because it appears relevant for the evolution of soils in the Petit Saint-Bernard Pass area, because it can help distinguish *e.g.*, Podzols from Stagnosols, as well as the different horizons of Podzols, and because it is a key constituent that is necessary for soil ecosystem services. Both axes are controlled by the soil-forming factors (Jenny, 1941), of which we emphasize the Human effect as a major soil-forming factor (Yaalon and Yaron, 1966; Richter, 2007). However, its presence as an individual factor is questionable because of its ability to change all the others. The third axis which accounts for parent material (p) and relief (r), represents the spatial dimension and the diversity of the pedo-landscape. The cycle begins with a phase of rapid soil growth (R), following glacial retreat and the formation of Leptosols and Cambisols, as described in the early Holocene soil evolution (Figs. 8 and 9). The soils reach a relative steady state, which is defined as the conservation phase (K). The organic carbon is maximal, and this phase corresponds to the Holocene climatic optimum and the maximal expansion of both Podzols (S_1) and *Pinus cembra* forest in the subalpine area (Figs. 8-d and 9). The Ω phase is initiated in our case by the beginning of human activities and represents a phase of release, characterized by deforestation, erosion and the haploidization of soil horizons. The haploidization process was introduced to describe the mixing and/or loss of soil horizons (Fig. 9) and contrasts with horizonation (Hole, 1961; Johnson and Watson-Stegner, 1987). Carbon stock decreases according to the horizons, and fluxes vary according to the intensity of the exported stock. Each single soil within the landscape follows a particular Ω trend that depends on local and temporal pressures, such as use as a campsite, overgrazing and according to the topography (the third dimension), which allows export and/or deposition of eroded material, thus multiplying the pedogenesis trajectories and increasing pedodiversity (Ibañez et al., 1995). The Ω phase is an intense, short and regressive phase that brings soils to a less developed state than they had during the K phase; and in some case to complete erosion of soils (S_0). The α phase constitutes a reorganization phase that is characterized by a rebalancing of the soil cover due to the existence of new environmental (*i.e.*, hydrology, vegetation, grazing…) conditions, leading to the development of Stagnosols (S_2) and to the beginning of a new cycle (Fig. 9). Contrary to (Holling, 1986), the α phase cannot be assimilated into a new R phase because soils do not return to their initial condition and develop from the degraded, previously existing soil. Yaalon and Yaron (1966) introduced the term of metapedogenesis (m_i) to indicate “the [hu]man-induced processes and changes in the soil profile”. The pristine soil evolution defines the pedogenesis, and the different Ω-α trajectories of the ecosystem define the anthropo-pedogenetic phases. Thus, the soil S_{i+1} (with $i \neq 0$) results from the anthropo-pedogenesis m_i applied to S_i and so on: $S_{i+1} = f(S_i, m_i)$. The Ω-α trajectories are directly involved in the provision of ecosystem services from the soil's natural capital (Dominati et al., 2010), and permit quantification and valuation of the ecosystem services provided by pasture soils (Dominati et al., 2014).

5. Conclusion

The study of the Holocene sediment sequence of Lake Verney and the analysis of current pedogenesis in the area of the Petit Saint-Bernard Pass allowed us to reconstruct the evolution of past soil processes in relation to past human activities. According to the substantial content of carbonate material in the sediment, decarbonatation seems to have been the main process that occurred during the first part of the Holocene. The acidification of the soils due to the presence of *Pinus cembra* vegetation and the increase in the Fe_{Ox}/Fe_{Tot} ratio until 6000 cal. yrs BP indicate the formation and migration of Fe organo-metallic compounds allowed the podzolization of the soils within the catchment during the mid-Holocene period. The beginning of human activities was recorded at 4300 cal. yrs BP, and was associated with land clearing and the successive erosion of the surface of Podzols horizons. During the Roman period, the spodic horizons were stripped away, meanwhile, erosion increased in response to sheep grazing. During the Middle Ages, pasturing way to cattle farming. Erosion increased again while the Fe_{Ox}/Fe_{Tot} ratio decreased, suggesting that the

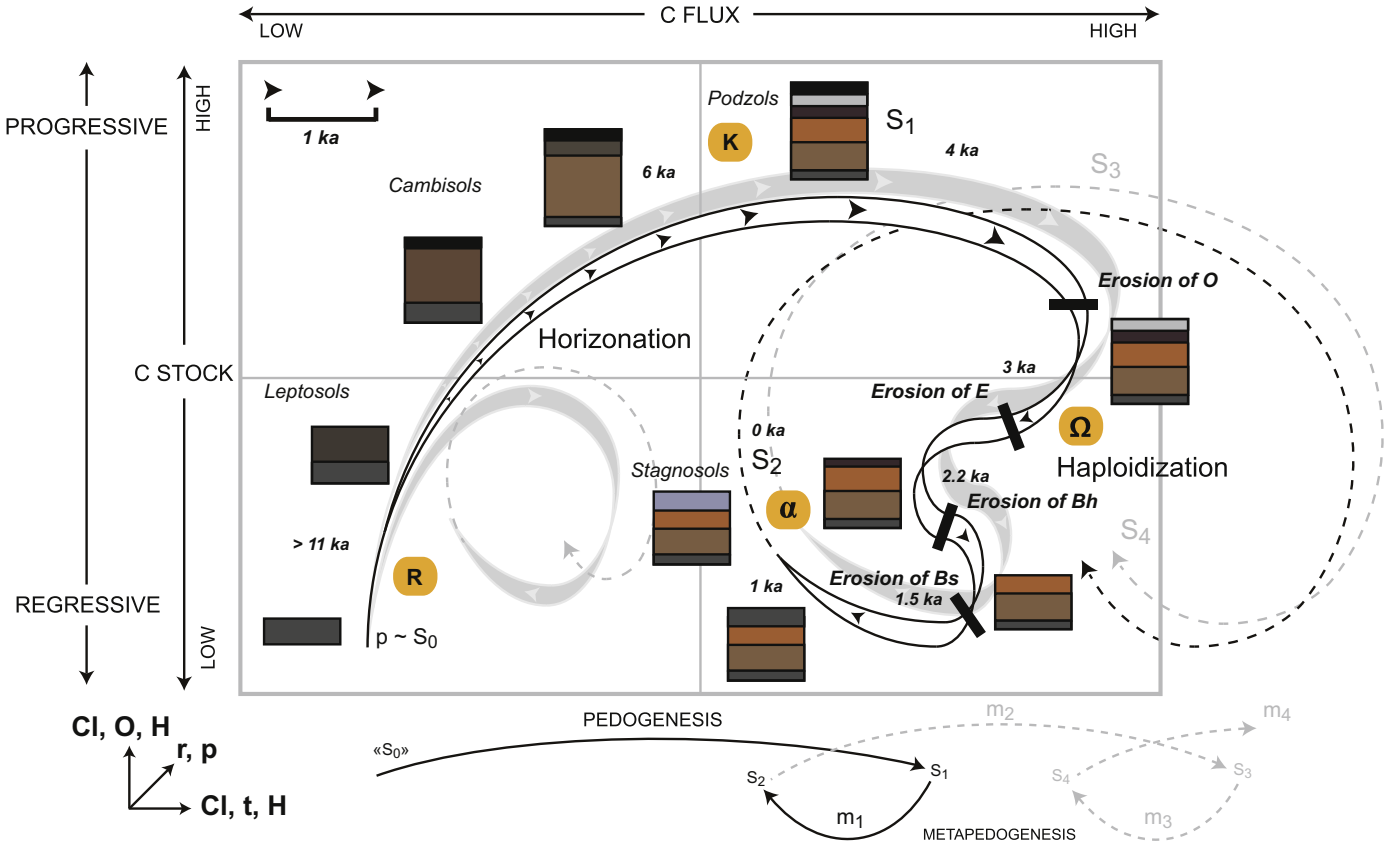


Fig. 9. Adaptation of the ecosystem evolution model of Holling (1986) to pedogenesis cycles and conceptual framework for man-made soil changes of Yaalon and Yaron (1966) applied to soil evolution in the Petit Saint Bernard Pass area. R corresponds to the rapid growth of soil, K to the conservation phase, Ω to the release phase and α to reorganization. Gray cycles in the background represent the diversity of the pedo-landscape according to the r and p factors (Jenny, 1941). Doted cycles represent future cycle trajectories. The intervals between arrowheads represent the speed, *i.e.*, a short interval means a slow change whereas a long interval indicates a rapid change. Abbreviations: C = carbon, Cl = climate, O = organism, H = human, r = relief, p = parent material, t = time, Si = soil i, mi = metapedogenetic phase i.

Podzols evolved into Stagnosols since that time. The study of paleosols, archaeological evidence and charcoal dating from soils in the area of the pass allowed validation of the soil evolution reconstruction from the lake sediment archive. The paleosols also confirmed the development of Stagnosols from degraded spodic horizons and provide additional information on erosional processes and the chronology of the Anthropocene. The combined studies of lake sediment archives and paleosols provide a long-term continuous reconstruction of pedogenesis at the landscape scale. Our study highlights the effect of the human soil-forming factor in the soil cycle evolution, and provides evidence of "anthropo-pedogenetic" phases. More than changes in vegetation and agro-pastoral activities, it is more the continuous use and progressive degradation of the soil cover that lead to soil changes. The change in soil type from Podzols to Stagnosols leads to reduced accumulation of organic matter that may have consequences for both the soil structure, favoring erosion, and carbon sequestration. This understanding of "release-reorganization" trajectories, *i.e.*, the anthropo-pedogenesis, could provide a possible means to quantify and value long-term soil ecosystem services.

Acknowledgments

The authors thank the city of La Thuile in Italy (Roberto Maddalena and Barbara Frigo) for access and coring authorizations and the ski resort of La Rosière for the winter access to the lake by the ski slopes. Thanks to Stéphanie Thiebault and Claire Delhon for charcoals identification of PSB1, Julie Deyrieux for the sedimentological analyses, Aurélien Borgeat, Manu Mallet and Pierre Faivre for field support. Thanks also to Sarah Bureau for ICP-AES analyses at ISTerre laboratory and to Delphine Rioux from LECA laboratory for DNA sampling and analyses. 14C analyses were acquired thanks to the CNRS-INSU ARTEMIS national radiocarbon AMS measurement program at Laboratoire de Mesure du 14C (LMC14) in the CEA Institute at Saclay (French Atomic Energy Commission). The authors thank also the Laboratoire Souterrain de Modane (LSM) facilities for the gamma spectrometry measurements. This research was performed by Hannibal Project financed by the DIPEE (Dispositifs de Partenariat en Ecologie et Environnement) Chambéry-Grenoble of the CNRS INEE.

References

Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services — a global review. Geoderma 262, 101–111. http://dx.doi.org/10.1016/j.geoderma.2015.08.009.
Anderson, H.A., Berrow, M.L., Farmer, V.C., Hepburn, A., Russell, J.D., Walker, A.D., 1982. A reassessment of podzol formation processes. Eur. J. Soil Sci. 33, 125–136.
Appleby, P., Oldfield, F., 1978. The calculation of lead-210 dates assuming a constant rate of supply of unsupported 210 Pb to the sediment. Catena 5, 1–8.
Appleby, P., Richardson, N., Nolan, P., 1991. 241Am dating of lake sediments. In: Environmental History and Palaeolimnology. Springer, pp. 35–42.
Arnaud, F., Revel-Rolland, M., Bosch, D., Winiarski, T., Desmet, M., Tribovillard, N., Givelet, N., 2004. A 300 year history of lead contamination in northern French Alps reconstructed from distant lake sediment records. J. Environ. Monit. 6, 448–456. http://dx.doi.org/10.1039/B314947A.
Arnaud, F., Révillon, S., Debret, M., Revel, M., Chapron, E., Jacob, J., Giguet-Covex, C., Poulenard, J., Magny, M., 2012. Lake Bourget regional erosion patterns reconstruction reveals Holocene NW European Alps soil evolution and paleohydrology. Quat. Sci. Rev. 51, 81–92. http://dx.doi.org/10.1016/j.quascirev.2012.07.025.
Arnaud, F., Poulenard, J., Giguet-Covex, C., Wilhelm, B., Révillon, S., Jenny, J.-P., Revel, M., Enters, D., Bajard, M., Fouinat, L., et al., 2016. Erosion under climate and human pressures: an alpine lake sediment perspective. Quat. Sci. Rev. 152, 1–18.
Baize, D., 2000. Guide des analyses en pédologie: 2e édition, revue et augmentée. (Editions Quae).
Bajard, M., Sabatier, P., David, F., Develle, A.-L., Reyss, J.-L., Fanget, B., Malet, E.,

Arnaud, D., Augustin, L., Crouzet, C., Poulenard, J., Arnaud, F., 2016. Erosion record in Lake La Thuile sediments (Prealps, France): evidence of montane landscape dynamics throughout the Holocene. The Holocene 26, 350–364.
Bajard, M., Poulenard, J., Sabatier, P., Develle, A.-L., Giguet-Covex, C., Jacob, J., Crouzet, C., David, F., Pignol, C., Arnaud, F., 2017. Progressive and regressive soil evolution phases in the Anthropocene. Catena 150, 39–52. http://dx.doi.org/10.1016/j.catena.2016.11.001.
Birkeland, P.W., 1999. Soils and Geomorphology, 3rd ed. Oxford University Press, New York.
Blaauw, M., 2010. Methods and code for “classical” age-modelling of radiocarbon sequences. Quat. Geochronol. 5, 512–518. http://dx.doi.org/10.1016/j.quageo.2010.01.002.
Bormann, B.T., Spaltenstein, H., McClellan, M.H., Ugolini, F.C., Cromack Jr., K., Nay, S.M., 1995. Rapid soil development after windthrow disturbance in pristine forests. J. Ecol. 747–757.
Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. Obitools: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Resour. 16, 176–182.
Brisset, E., Miramont, C., Guiter, F., Anthony, E.J., Tachikawa, K., Poulenard, J., Arnaud, F., Delhon, C., Meunier, J.-D., Bard, E., Sumera, F., 2013. Non-reversible geosystem destabilisation at 4200 cal. BP: sedimentological, geochemical and botanical markers of soil erosion recorded in a Mediterranean alpine lake. The Holocene 23, 1863–1874. http://dx.doi.org/10.1177/0959683613508158.
Brisset, E., Guiter, F., Miramont, C., Troussier, T., Sabatier, P., Poher, Y., Cartier, R., Arnaud, F., Malet, E., Anthony, E.J., 2017. The overlooked human influence in historic and prehistoric floods in the European Alps. Geology G38498–1.
COM EU, 2006. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - Thematic Strategy for Soil Protection [SEC(2006)620] [SEC(2006)1165]/* COM/2006/0231 Final*/.
Crogiez-Petrequin, S., 2006. Col du Petit Saint Bernard: Les fouilles du bâtiment ouest 2003–2005 Epoque Gallo-Romaine, in: Actes Du Séminaire de Clôture Du Programme INTERREG III A ALCOTRA 2000–2006: Alpis Graia. Archéologie sans Frontière Au Col Du Petit St-Bernard. Aoste. pp. 131–141.
Daily, G.C., Matson, P.A., Vitousek, P.M., 1997. Ecosystem services supplied by soil. In: Nature's Services: Societal Dependence on Natural Ecosystems, pp. 113–132.
Dambrine, É., 1985. Contribution à l'étude de la répartition et du fonctionnement des sols de haute montagne: massifs des Aiguilles Rouges et du Mont-Blanc.
David, F., 1993. Évolutions de la limite supérieure des arbres dans les Alpes françaises du Nord depuis la fin des temps glaciaires.
de Beaulieu, J., Kostenzer, J., Reich, K., 1993. Dynamique forestière holocène dans la haute vallée de l'Arve (Haute-Savoie) et migrations de Abies et Picea dans les Alpes occidentales. Dissertationes Botanicae. 196. pp. 387–398.
Dearing, J., Hu, Y., Doody, P., James, P.A., Brauer, A., 2001. Preliminary reconstruction of sediment-source linkages for the past 6000 yr at the Petit Lac d'Annecy, France, based on mineral magnetic data. J. Paleolimnol. 25, 245–258.
Development Core Team, R., 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol. Econ. 69, 1858–1868. http://dx.doi.org/10.1016/j.ecolecon.2010.05.002.
Dominati, E., Mackay, A., Green, S., Patterson, M., 2014. A soil change-based methodology for the quantification and valuation of ecosystem services from agro-ecosystems: a case study of pastoral agriculture in New Zealand. Ecol. Econ. 100, 119–129. http://dx.doi.org/10.1016/j.ecolecon.2014.02.008.
Doyen, E., Etienne, D., 2017. Ecological and human land-use indicator value of fungal spore morphotypes and assemblages. Veg. Hist. Archaeobotany 26, 357–367. http://dx.doi.org/10.1007/s00334-016-0599-2.
Dreibrodt, S., Wiethold, J., 2015. Lake Belau and its catchment (northern Germany): a key archive of environmental history in northern central Europe since the onset of agriculture. The Holocene 25, 296–322.
Elbaz-Poulichet, F., Dezileau, L., Freydier, R., Cossa, D., Sabatier, P., 2011. A 3500-year record of hg and Pb contamination in a Mediterranean sedimentary archive (the Pierre Blanche Lagoon, France). Environ. Sci. Technol. 45, 8642–8647. http://dx.doi.org/10.1021/es2004599.
Enters, D., Dorfler, W., Zolitschka, B., 2008. Historical soil erosion and land-use change during the last two millennia recorded in lake sediments of Frickenhauser See, northern Bavaria, central Germany. The Holocene 18, 243–254. http://dx.doi.org/10.1177/0959683607086762.
Etienne, D., Jouffroy-Bapicot, I., 2014. Optimal counting limit for fungal spore abundance estimation using Sporormiella as a case study. Veg. Hist. Archaeobotany 23, 743–749. http://dx.doi.org/10.1007/s00334-014-0439-1.
Etienne, D., Wilhelm, B., Sabatier, P., Reyss, J.-L., Arnaud, F., 2013. Influence of sample location and livestock numbers on Sporormiella concentrations and accumulation rates in surface sediments of Lake Allos, French Alps. J. Paleolimnol. 49, 117–127. http://dx.doi.org/10.1007/s10933-012-9646-x.
Faegri, K., Iversen, J., 1989. Textbook of Pollen Analysis, Revised by Faegri K, Kaland PE, Krzywinski K. J Wiley, New York.
FAO, 2006. Guidelines for Soil Description, 4, [rev.] ed. ed. FAO, Rome.
Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., et al., 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol. Ecol. Resour. 15, 543–556.
Giguet-Covex, C., Arnaud, F., Poulenard, J., Disnar, J.-R., Delhon, C., Francus, P., David, F., Enters, D., Rey, P.-J., Delannoy, J.-J., 2011. Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (Lake Anterne, 2063 m a.s.l., NW French Alps): the role of climate and human activities. The Holocene 21, 651–665. http://dx.doi.org/10.1177/0959683610391320.
Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J., Taberlet, P., 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat. Commun. 5. http://dx.doi.org/10.1038/ncomms4211.
Goldberg, E.D., 1963. Geochronology with 210Pb. In: Radioactive Dating, pp. 121–131.
Heiri, O., Lotter, A.F., Lemcke, G., 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. J. Paleolimnol. 25, 101–110.
Hole, F.D., 1961. A classification of pedoturbations and some other processes and factors of soil formation in relation to isotropism and anisotropism. Soil Sci. 91, 375–377.
Holling, C.S., 1986. The resilience of terrestrial ecosystems: local surprise and global change. In: Sustainable Development of the Biosphere. 14 (292À317).
Hošek, J., Pokorný, P., Prach, J., Lisá, L., Grygar, T.M., Knésl, I., Trubač, J., 2017. Late Glacial erosion and pedogenesis dynamics: evidence from high-resolution lacustrine archives and paleosols in south Bohemia (Czech Republic). Catena 150, 261–278. http://dx.doi.org/10.1016/j.catena.2016.11.022.
Ibañez, J.J., De-Albs, S., Bermúdez, F.F., García-Álvarez, A., 1995. Pedodiversity: concepts and measures. Catena 24, 215–232.
Jeanroy, E., 1983. Diagnostic des formes du fer dans les pédogénèses tempérées: évaluation par les réactifs chimiques d'extraction et apports de la spectrométrie Mossbauer. Université de Nancy I.
Jenny, H., 1941. Factors of Soil Formation—-a Sytem of Quantitative Pedology. McGraw-Hill, New York.
Johnson, D.L., Watson-Stegner, D., 1987. Evolution model of pedogenesis. Soil Sci. 143, 349–366.
Krishnaswamy, S., Lal, D., Martin, J.M., Meybeck, M., 1971. Geochronology of lake sediments. Earth Planet. Sci. Lett. 11, 407–414. http://dx.doi.org/10.1016/0012-821X(71)90202-0.
Larsen, I.J., Almond, P.C., Eger, A., Stone, J.O., Montgomery, D.R., Malcolm, B., 2014. Rapid soil production and weathering in the southern Alps, New Zealand. Science 343, 637–640. http://dx.doi.org/10.1126/science.1244908.
Le Quéré, C., Andrew, R.M., Canadell, J.G., Sitch, S., Korsbakken, J.I., Peters, G.P., Manning, A.C., Boden, T.A., Tans, P.P., Houghton, R.A., Keeling, R.F., Alin, S., Andrews, O.D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L.P., Ciais, P., Currie, K., Delire, C., Doney, S.C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Klein Goldewijk, K., Jain, A.K., Kato, E., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Melton, J.R., Metzl, N., Millero, F., Monteiro, P.M.S., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S., O'Brien, K., Olsen, A., Omar, A.M., Ono, T., Pierrot, D., Poulter, B., Rödenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B.D., Sutton, A.J., Takahashi, T., Tian, H., Tilbrook, B., van der Laan-Luijkx, I.T., van der Werf, G.R., Viovy, N., Walker, A.P., Wiltshire, A.J., Zaehle, S., 2016. Global carbon budget 2016. Earth Syst. Sci. Data 8, 605–649. http://dx.doi.org/10.5194/essd-8-605-2016.
Legros, J.-P., 2007. Les grands sols du monde. PPUR presses polytechniques.
Legros, J., Martini, I., 1992. Soils of Alpine mountains. In: Weathering, Soils and Paleosols, pp. 155–181.
Leveau, P., Segard, M., 2006. Le pastoralisme antique autour du col du Petit-Saint-Bernard. Alpis Graia, Archéologie sans frontières au col du Petit-Saint-Bernard, Seminario di chiusura, Aosta, 2–4 marzo 2006 153–161.
Malkiewicz, M., Waroszewski, J., Bojko, O., Egli, M., Kabala, C., 2016. Holocene vegetation history and soil development reflected in the lake sediments of the Karkonosze Mountains (Poland). The Holocene, 0959683615622546.
Massa, C., Bichet, V., Gauthier, É., Perren, B.B., Mathieu, O., Petit, C., Monna, F., Giraudeau, J., Losno, R., Richard, H., 2012. A 2500 year record of natural and anthropogenic soil erosion in South Greenland. Quat. Sci. Rev. 32, 119–130. http://dx.doi.org/10.1016/j.quascirev.2011.11.014.
McKeague, J., Brydon, J.E., Miles, N.M., 1971. Differentiation of forms of extractable iron and aluminum in soils. Soil Sci. Soc. Am. J. 35, 33–38.
Millet, L., Ortu, E., Miras, Y., Heiri, O., 2008. Les assemblages de chironomidae et les cortèges polliniques, outils de reconstitutions quantitatives des changements du climat Holocène: une approche croisée au col du Petit Saint Bernard.
Miras, Y., Millet, L., Guiter, F., Ponel, P., De Beaulieu, J.-L., Gozlar, T., 2006. Dynamique des écosystèmes et impact de l'homme dans le secteur du col du Petit Saint Bernard au cours de l'Holocène, in: Actes Du Séminaire de Clôture Du Programme INTERREG III A ALCOTRA 2000–2006: Alpis Graia. Archéologie sans Frontière Au Col Du Petit St-Bernard. Aoste. pp. 31–50.
Montagne, D., Cousin, I., Cornu, S., 2016. Changes in the pathway and the intensity of albic material genesis: role of agricultural practices. Geoderma 268, 156–164. http://dx.doi.org/10.1016/j.geoderma.2016.01.019.
Montgomery, D.R., 2007. Soil erosion and agricultural sustainability. Proc. Natl. Acad. Sci. 104, 13268–13272.
Moulin, B., Rey, P.-J., 2008. Les séquences pédo-sédimentaires des versants du col du Petit Saint Bernard. In: Table Ronde Du GDR JURALP. COLLECTION EDYTEM - Cahiers de Paléoenvironnement, Aix-en-Provence, France, pp. 191–206.
Mourier, B., 2008. Contribution de l'approche sédimentologique à la reconstitution de l'histoire des sols. In: Définition de traceurs pédologiques et application sur des sédiments lacustres de montagne (Maurienne, Savoie, France). Université de Savoie.
Mourier, B., Poulenard, J., Chauvel, C., Faivre, P., Carcaillet, C., 2008. Distinguishing subalpine soil types using extractible Al and Fe fractions and REE geochemistry. Geoderma 145, 107–120. http://dx.doi.org/10.1016/j.geoderma.2008.03.001.
Mourier, B., Poulenard, J., Carcaillet, C., Williamson, D., 2010. Soil evolution and subalpine ecosystem changes in the French Alps inferred from geochemical analysis of

lacustrine sediments. J. Paleolimnol. 44, 571–587. http://dx.doi.org/10.1007/s10933-010-9438-0.
Oldeman, L., 1994. The Global Extent of Soil Degradation. Soil Resilience and Sustainable Land Use. 9.
Pallmann, H., 1947. Pédologie et Phytosociologie. Comptes rendus de la Conférence de Pédologie Méditerranéenne. Springerpp. 3–36 (Editions AFES).
Pansu, J., Giguet-Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., Choler, P., 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Mol. Ecol. 24, 1485–1498.
Poulenard, J., Arnaud, F., Perrette, Y., Sabatier, P., Deline, P., Giguet-Covex, C., Bajard, M., Mourier, B., Quiers, M., Malet, E., Develle, A.-L., Fanget, B., Faivre, P., Dambrine, É., Chalmin, E., 2015. Organic geochemistry of soils and sediments: a key to describe the anthropocene? In: Collection EDYTEM - Sols et Matières Organiques: Mémoires et Héritages. Presented at the Sessions communes des 12ème journées d'Etude de Sols - Sols en héritages et de la 2ème réunion des chercheurs francophones en géochimie organique, Laboratoire EDYTEM, Université de Savoie 30 juin - 4 juillet 2014, pp. 9–18.
Protz, R., MARTINI, I., Ross, G.J., Terasmae, J., 1984. Rate of podzolic soil formation near Hudson Bay, Ontario. Can. J. Soil Sci. 64, 31–49.
Puydarrieux, P., Kervinio, Y., Darses, O., 2016. EFESE - Evaluation Française des Ecosystèmes et des Services Ecosystémiques - Rapport intermédiaire.
Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years Cal BP.
Rémy, B., Lempereur, O., Amandry, M., 2006. Les monnaies Antiques découvertes au col du Petit SAint Bernard (versants Français et Italien), in: Actes Du Séminaire de Clôture Du Programme INTERREG III A ALCOTRA 2000–2006: Alpis Graia. Archéologie sans Frontière Au Col Du Petit St-Bernard. Aoste. pp. 163–179.
Rey, P.-J., 2011. Réservoirs et systèmes d'irrigation dans les alpages du col du Petit Saint Bernard: vers l'identification de structures antiques. Les Cahiers du CRHIPA. pp. 354.
Rey, P.-J., Moulin, B., 2006. Occupations et circulations pré-romaines autour du Col du Petit-Saint-Bernard; méthode et premiers résultats d'une étude archéologique et sédimentaire de la montagne alpine. In: Projet Interreg IIIA, Seminario Di Chiusura. Musumeci SPA, pp. 77–118.
Rey, P.-J., Batigne-Vallet, C., Collombet, J., Delhon, C., Martin, L., Moulin, B., Oberlin, C., Poulenard, J., Robin, V., Thiebault, S., Treffort, J.-M., 2014. Approche d'un territoire de montagne: occupations humaines et contexte pédo-sédimentaire des versants du col du Petit-Saint-Bernard, de la Préhistoire à l'Antiquité. In: Senépart, I., Billiard, C., Bostyn, F., Praud, Y., Thirault, E. (Eds.), Méthodologie des recherches de terrain sur la Préhistoire récente en France. Nouveaux acquis, nouveaux outils, 1987–2012. Actes des premières Rencontres Nord-Sud de Préhistoire récente, Marseille, mai 2012. 73–71.
Rey, P.-J., Franc, O., Fudral, S., Moulin, B., Moulin, 2015. Le Cercle de pierres dressés du Col du Petit Saint-Bernard (Savoie - Val d'Aoste, 2188 m d'altitude) Nouvelles données de terrain et pistes d'interprétation. Société Valdôtaine de préhistoire et d'archéologie. pp. 163–190.
Reyss, J.-L., Schmidt, S., Legeleux, F., Bonté, P., 1995. Large, low background well-type detectors for measurements of environmental radioactivity. In: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 357. pp. 391–397.
Richter, D., 2007. Humanity's transformation of Earth's soil: Pedology's new frontier. Soil Sci. 172, 957–967.
Robbins, J.A., Edgington, D.N., 1975. Determination of recent sedimentation rates in Lake Michigan using Pb-210 and Cs-137. Geochim. Cosmochim. Acta 39, 285–304.
Sabatier, P., Dezileau, L., Briqueu, L., Colin, C., Siani, G., 2010. Clay minerals and geochemistry record from northwest Mediterranean coastal lagoon sequence: implications for paleostorm reconstruction. Sediment. Geol. 228, 205–217. http://dx.doi.org/10.1016/j.sedgeo.2010.04.012.
Sabatier, P., Poulenard, J., Fanget, B., Reyss, J.-L., Develle, A.-L., Wilhelm, B., Ployon, E., Pignol, C., Naffrechoux, E., Dorioz, J.-M., et al., 2014. Long-term relationships among pesticide applications, mobility, and soil erosion in a vineyard watershed. Proc. Natl. Acad. Sci. 111, 15647–15652.
Shand, C.A., Wendler, R., 2014. Portable X-ray fluorescence analysis of mineral and organic soils and the influence of organic matter. J. Geochem. Explor. 143, 31–42. http://dx.doi.org/10.1016/j.gexplo.2014.03.005.
Simonneau, A., Doyen, E., Chapron, E., Millet, L., Vannière, B., Di Giovanni, C., Bossard, N., Tachikawa, K., Bard, E., Albéric, P., Desmet, M., Roux, G., Lajeunesse, P., Berger, J.F., Arnaud, F., 2013. Holocene land-use evolution and associated soil erosion in the French Prealps inferred from Lake Paladru sediments and archaeological evidences. J. Archaeol. Sci. 40, 1636–1645. http://dx.doi.org/10.1016/j.jas.2012.12.002.
Stockmarr, J., 1971. Tablets with spores used in absolute pollen analysis. Pollen Spores 13, 615–621.
Stucki, J.W., Goodman, B.A., Schwertmann, U., 1988. Iron in Soils and Clay Minerals.
Taberlet, P., PRUD'HOMME, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., et al., 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Mol. Ecol. 21, 1816–1820.
Talon, B., 2006. Analyses anthracologiques au col du Petit St-Bernard. Archéoanthracologie et pédoanthracologie., in: Actes Du Séminaire de Clôture Du Programme INTERREG III A ALCOTRA 2000–2006: Alpis Graia. Archéologie sans Frontière Au Col Du Petit St-Bernard. Aoste. pp. 51–58.
Tamm, O., 1922. Urn bestamning ow de organisk komponenterna i merkens gelkomplex. Medd. Slatens. Skogsforsokanst. 19. pp. 384.
Trimble, S.W., Mendel, A.C., 1995. The cow as a geomorphic agent—a critical review. Geomorphology 13, 233–253.
Vanwalleghem, T., Gómez, J.A., Infante Amate, J., González de Molina, M., Vanderlinden, K., Guzmán, G., Laguna, A., Giráldez, J.V., 2017. Impact of historical land use and soil management change on soil erosion and agricultural sustainability during the Anthropocene. Anthropocene 17, 13–29. http://dx.doi.org/10.1016/j.ancene.2017.01.002.
Willis, K.J., Braun, M., Sümegi, P., Tóth, A., 1997. Does soil change cause vegetation change or vice versa? A temporal perspective from Hungary. Ecology 78, 740–750.
Wilmshurst, J.M., McGlone, M.S., 2005. Origin of pollen and spores in surface lake sediments: comparison of modern palynomorph assemblages in moss cushions, surface soils and surface lake sediments. Rev. Palaeobot. Palynol. 136, 1–15. http://dx.doi.org/10.1016/j.revpalbo.2005.03.007.
WRB - FAO, 2014. World Reference Base for Soil Resources 2014 International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome.
Yaalon, D.H., Yaron, B., 1966. Framework for man-made soil changes-an outline of metapedogenesis. Soil Sci. 102, 272–277.