



**Dipartimento di Scienze e Tecnologie per l'Agricoltura,
le Foreste, la Natura e l'Energia - DAFNE**

Corso di Dottorato di ricerca in
Scienze e tecnologie per la gestione forestale e ambientale - XXVIII ciclo

Spatiotemporal analysis and modeling of ecological processes at ecosystem, landscape and bioregion scale

**Analisi spaziale e modellizzazione spaziotemporale di
processi a scala di ecosistema, paesaggio e bioregione
(AGR/10)**

Tesi di dottorato di **Marco Ciolfi**

Tutor: **Prof. Gianluca Piovesan**

firma.....

firma.....

Coordinatore **Prof. Bartolomeo Schirone**

Co-tutor: **Dr. Marco Lauteri** CNR-IBAF

firma.....

firma.....

Spatiotemporal analysis and modeling of ecological processes at ecosystem, landscape and bioregion scale

TimescapeGlobal software release notes

The development of **TimescapeGlobal** software was part of the author's activities within his Doctorate project under the supervision of Prof. Gianluca Piovesan of the Università degli studi della Toscana, funded by the National Research Council of Italy grant IBAF003/2013/POR, under the direction of Dr. Marco Lauteri of the Institute of Agro-Environmental and Forest Biology [IBAF].

The development of **TimescapeGlobal** wouldn't have been possible without the insightful discussions with the *Spatio-temporal Isotope Analytics Lab* of the University of Utah [SPATIAL]. The author is particularly grateful to Prof. Gabriel Bowen and Dr. Annie Putman, who also provided the sample datasets included in the distribution; the mission was funded by the 2015 Short Term Mobility Program of the Department of Earth systems science and environmental technologies of the National Research Council of Italy, grant no.0051519. The author is grateful as well to Prof. Cristina Máguas of the *Centro de Biologia Ambiental* of the Faculty of Sciences of the University of Lisbon [CBA] for her helpful advice in stable isotopes ecology and spatial modeling; the mission was funded by COST short term scientific mission, grant no.TD1209-310314-043127.

TimescapeGlobal is free software, it is released according to the GNU GPLv3 License, see [GNU] for details. The source code is part of the standard distribution. Users are allowed and encouraged to modify any part of the software according to their particular needs. Users are also encouraged to publish their contributions following the same licensing policy. Users cannot include parts of this software into non-open derivative products.

software download URI: <https://sourceforge.net/projects/timescapeglobal/>



Contents

0	Introduction	3
1	State of the Art in Space and Time Modeling	9
1.1	Spatial Modeling	9
1.2	Temporal Modeling	11
1.3	Spatiotemporal Modeling	12
2	The Timescape Algorithm	14
2.1	Evaluation of Distances	16
2.2	Causal Constraints	18
2.3	Model Evaluation	21
2.3.1	Model Tuning	22
2.4	Model Storage and Exploration	24
3	The Software Packages	25
3.1	TimescapeLocal	26
3.1.1	The MANAGER Panel	26
3.1.2	The SETUP Panel	27
3.1.3	Model Interpolation	30
3.1.4	The EXPLORATION Panel	32
3.2	TimescapeGlobal	40
3.2.1	The SETUP Panel	40
3.2.2	The Model Evaluation	43
3.2.3	The EXPLORATION Panel	45
4	Case Study: Mycorrhiza Survival Strategy	49
4.1	Description of the Area	49
4.2	The Symbiosis Model	52
4.3	Materials and Methods	53
4.3.1	Sampling	53
4.3.2	Isotopic Analyses	54
4.3.3	Spatial Statistics	54
4.4	Results	56
4.4.1	Ordinary Geostatistical Analysis	56

4.4.2	Timescape Analysis	57
4.5	Discussion	60
4.5.1	Conclusion: Complex Modeling for Complex Environments	62
5	Case Study: Olive Oil Provenance Assessment	63
5.1	EVOO Isotopic Content	63
5.2	Materials and Methods	64
5.2.1	Geospatial Model	66
5.3	Results	67
5.3.1	Predicted ^{18}O Isoscape	70
5.3.2	Predicted ^{13}C Isoscape	71
5.3.3	Conclusion: Protection from <i>geofrauds</i>	72
6	Case Study: Space-time Sampling Planning	73
6.1	The Project	74
6.2	The Sampling Planning and Campaign	75
7	Conclusions and Outlook	78
A	Appendix: The Timescape Maths	80
A.1	Space and Time Distances	80
A.2	Causal Structure	83
A.3	Building the Model	86
A.4	Evaluation Algorithm	88
A.5	Jackknifing and Ensemble Means	90
B	Appendix: The Timescape Data Structure	93
C	Appendix: Mycorrhiza Survival Strategy Sampling Atlas	97
D	Appendix: Mycorrhiza Survival Strategy Statistics	110
	References	123
	Index	133

0 Introduction

From a statistical point of view, ecological systems are very complex: many variables cooperate in the definition of the sampled values and many disturbances affect the sampling. Unlike a lab-based experiment, where one keeps the independent variables under control, field sampling brings any kind of uncertainty into the collected data. Furthermore, ecological systems like the forest environments often show patterns of change both in time and space.

Complexity calls for compromise: choosing a single driver of change, be it space or time, is always somewhat arbitrary and the researchers are often forced to employ pre-packaged statistical analysis tools for their needs. Sometimes, however, it is hard to spot the main source of change between space and time; they both contribute on similar grounds. While established tools do exist for geostatistics and time series analysis as well, there is a lack of tools whenever it is not possible to choose a primary source of variability, neglecting the other.

This work proposes an innovative technique to treat space and time variability at the same time, especially tailored for ecological systems. The algorithm, borrowed from statistical physics and field theory, is adapted to the complexity of the natural environment, it can be applied to a variety of measurements in the field of forest ecology and other Earth and environmental sciences. As of now, applications have been in the field of stable isotopes.

Aside from the results already obtained, the proposed interpolation technique looks promising for many contemporary fields of ecology research related to climate change issues, where the temporal patterns play a key role.

Space and Time Variability in Forest Ecosystems

Transport and diffusion phenomena take place always and everywhere in natural environments, at any scale. The forest systems offers a great statistical opportunity: the presence of *natural data recorders*, mainly trees, that store informations about the past and present conditions into their tissues. This is especially true in the field of the stable isotopes, which can be used as ecophysiological markers, as is the case of photosynthetic fractionation of the carbon isotopes [Brugnoli and Farquhar 2000]. All tissues build up over time getting their matter content from the surrounding environment, thus becoming *information integrators*. Sometimes (e.g. tree rings) a single sample is actually a time series of observations of one or more variables. Furthermore, the widespread use of the now cheap GPS technology gives us the opportunity of a fast and reliable placement of the samples into a space frame. This all happens at a variety of scales, from a single forest plot to the whole Earth.

Geostatistics has entered the realm of forest ecology long ago [Rossi et al. 1992]. There are issues about the methodological correctness of the application of space interpolation techniques to the kind of datasets normally used in forest research¹ but, as is customary in environmental sciences, there is a certain consensus for a loose application of such algorithms, safe an a posteriori cross-check with a suitable control group.

The sources of variability can be both internal or external to the ecosystem. The internal sources include all the occurring biochemical processes, while the external sources are related to the changing environmental conditions, and to the energy input and output (e.g. the sunlight in photosynthesis): forests are not closed systems, thermodynamically speaking. A forest is a ever-changing collection of already complex individuals, connected by an intricate network of relationships; these relationships occur both in space and in time and neglecting a driver of variability is always arbitrary.

The Timescape Idea

The leading idea of the Timescape algorithm is twofold: on the one hand time is transformed into a third spatial dimension, on the other hand a *causal structure* is imposed in order to incorporate an *evolutionary pattern* into the models. This is just the mathematical translation of a few common sense statements:

- the value of a certain quantity² at a point in space is influenced by the past conditions in the same place,
- the value at a given time can influence the future values,
- closer points can influence each other more than far ones,³
- observations at different sites and times can be mixed freely,
- the area of influence of a point grows with time.

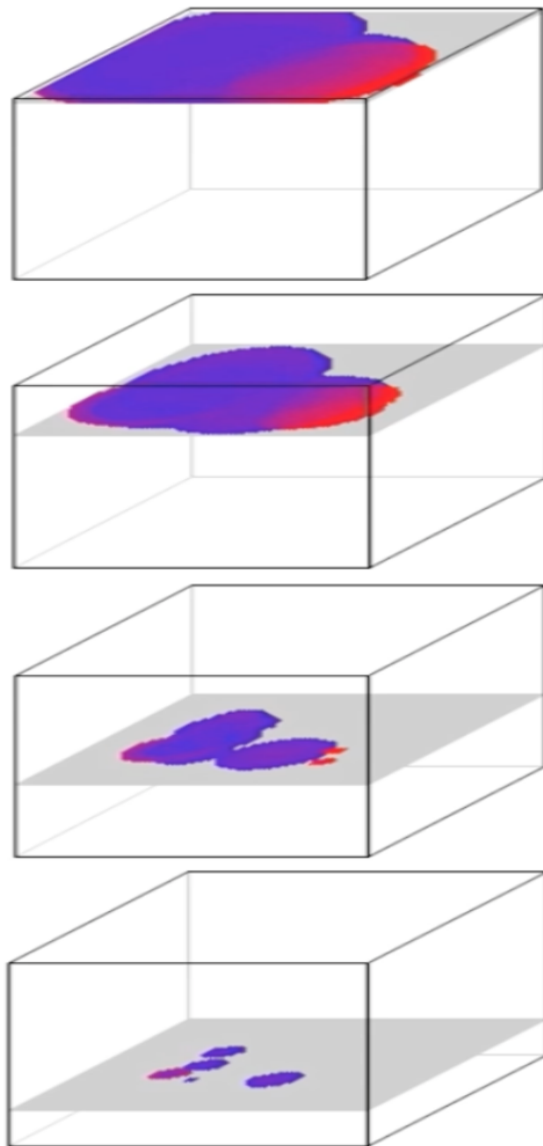
In a single sentence: the values investigated are treated as actual objects that “flow”, diffusing from point to point like a drop of ink into a volume of water. The samples dataset is the collection of drops that feed this flow. It is the information about the values which actually flows, however, not a physical object.

¹Major problems are the *stability* and *non-autocorrelation* of datasets, a constraint which is required, as a matter of principle, but almost never met [Cressie 1990].

²By *quantity* we mean any measure that outputs a number. Statisticians say *random variable*.

³In a nutshell, this is the foundation of geostatistics [Fortin and Dale 2011].

The idea is nothing new in field theory and statistical mechanics, but the application to ecological systems adds some difficulties. First of all, forests are not lab-tailored systems, there are countless sources of noise and the correlations are always approximate. This is also the case of ordinary geostatistical methods, of course, so it is not a major concern here. The goal is to exploit the time variability to consolidate the dataset, rather than flattening it, neglecting the differences in observation time and enlarging the errors in the outcomes.



A *Timescape*, sliced according to increasing time (bottom to top). Every single slice is a full spatial model.

The conversion of the time into a third spatial dimension serves a definite strategy: it allows the use of geostatistical methods without concerns about the methodological correctness, which is inherited almost automatically from the far more common spatial-only techniques. Timescape is not more questionable than any common geostatistical interpolation technique (of course, it is not better than the other methods, too).

A major obstacle towards the treatment of three (or more) dimensional models is the need for a large memory space, both for calculations and for storage. Up to a few years ago, it was the realm of supercomputing but in recent years the availability of affordable hardware at relatively cheap prices allowed the implementation of complex algorithms with a regular desktop-based system.⁴ A discrete three-dimensional model can be composed of billions of cells, so a database management system is mandatory for the storage.

The R statistical system [R] offers a stable, robust and friendly environment for any sort of statistical calculations, it is also widespread in the field of forest ecology [Bivand et al. 2008, Borcard et al. 2011, Zuur et al. 2007] but R needs to store all the data into the computer's RAM for calculations, so it is not the right choice for such big models. For these reasons and for portability the Algorithm was developed in Java language, using an external Oracle or MySQL Database [JAVA, ORACLE, MYSQL].

What Timescape is for

The Timescape algorithm (and the actual software package) has been developed for the typical issues of forest ecology studies, however, it can be useful in many other fields of application, mainly concerning Earth sciences, at various spatial scales. Extensive testing has been conducted on precipitation waters stable isotopes ratios [GNIP] at a continental to worldwide scale.

Forest environments bring an added complexity, from a geostatistical point of view, since they often show a multi-scale variability which threatens some of the data stability requirements of many spatial statistical techniques. Sometimes, adding the time component as a third independent variable can improve stability; for example, if there is an harmonic (say periodic or seasonal) component of variability, as it happens frequently as far as plant growth is concerned, the time coordinate splits away the clustered samples.

For these reasons, the range of Timescape extends well beyond the forest ecology, including many other field, from biology to Earth sciences, if an evolutionary phenomenon has to be investigated. This is especially useful whenever periodic changes over time are involved.

The Timescape development has been done with the broadest possible range of computers in mind: the calculation strategy has been shaped around a standard hardware. The actual calculation times can be as long as a few days, but it is only a matter of patience and model resolution. Downscaled models can be estimated in as few as half an hour, before moving

⁴The development of **Timescape** started on an dual-i7, 8GB MacAir laptop. The project than moved on a 3.5 GHz 6-XeonE5 64GB MacPro desktop system.

to bulkier ones. The Algorithm is public, it is not subject to any registration whatsoever; the software package is published with an open license and the software needs no third-party expensive components to run. The adhesion to a clean open development philosophy brings a twofold benefit for a scientific application: firstly there are no licensing fees and, far more important, any researcher can know exactly what his/her tool is doing, up to the last line of code, which is especially important from a methodological of view.

The development of the Timescape Algorithm followed a *minimum detour path* from a supposedly standard GIS⁵ workflow in forest ecology studies. This minimum departure is chosen according to the constraint of an as-standard-as-possible approach.

What Timescape is not for

Timescape is not by far the standard tool for any interpolation. It is a discrete model based on the supposed independence of one cell from another, the values interpolated are only a function of the distribution of the samples dataset.⁶ This makes the tool unsuitable for all problems with a strong inter-cell connection, e.g. anything concerning meteorology.

The Timescape algorithm is *blind*: it makes no assumptions about an underlying evolutionary model. If it is the case, when the modelled phenomenon follows a known pattern,⁷ it is much better to follow a different approach, through the numerical modeling of the equations. This is almost never the case in forestry and in complex ecological systems in general; this is not to say that cause-effect relationships do not exist, but that they are too complex to be represented with an equation or a set of equations.

The use of Timescape is recommended in all the cases when the user is in trouble deciding whether the primary source of variability is time or displacement; its use is disproportionate when the observation are clustered in time, i.e. if one has an homogeneous sampling repeated from season to season: in this case the results of Timescape essentially overlap those of ordinary spatial interpolators but the cost in time of computing is much higher.

Lastly, the current implementation of Timescape approximates the Earth with a sphere: it cannot be used if the subtleties of the actual Earth shape are involved, as in geophysical modeling.

⁵Geographical Information Systems (GIS) are widespread in the field of forest ecology. They are often the ideal choice for data storage and elaboration.

⁶It is a standard assumption of most of the spatial statistics ordinary algorithms, as Kriging or IDW (Inverse Distance Weighting).

⁷i.e. when a differential equation exists which accounts for the observations values.

Plan of the Disstertation

The Timescape Algorithm was conceived within a study of symbiosis among *Tuber aestivum*, pines and maybe oaks and other trees. The symbiosis was modelled with the ^{13}C and ^{15}N fractionation between host and mycorrhiza, building spatial models of isotopic relative abundances, the so-called *Isoscapes* (chapter 4). The main difficulties arose for the continuous nitrogen fractionation during the months of collection, giving rise to comparatively too large errors with the traditional spatial interpolation techniques.

Chapter 1 describes briefly the state of the art of space and time modeling in complex ecosystems. The Timescape algorithm is presented in chapter 2, while the resulting software is presented in detail in chapter 3. The complete manual of the software is included in the distribution package [TIMESCAPEGLOBAL]. Appendix A is a deep technical discussion of all the subtleties of the algorithm. Appendix B describes the Database structure.

Three case studies illustrate practical space-time modeling in forest ecology:

- Chapter 4, as said, shows the Timescape algorithm at work in an investigation about symbiotic relationships through stable isotopes measurements. This study involved the analysis of more than one thousand samples, just on the isotopic side, thus providing an ideal, spatially complete playground for geostatistical modeling.⁸ Chapter 3, in fact, illustrates in detail all the building steps of the Timescape model involved in this study.
- Chapter 5 illustrates an example of geostatistical modeling, not involving Timescapes, dealing with the geographical assessment of Italian extra virgin olive oils, for the prevention of frauds in the high-quality oil market.⁹
- Chapter 6 describes a real-life geostatistical sampling plan in an almost unknown forest environment: the isotopic sampling of *Polylepis reticulata* within the Ecuaflex project in the Ecuadorean Andes.

Chapter 7 outlines the possible evolution of spacetime modeling in the field of forest ecology.

⁸The software has been presented at the 2016 European Geospatial Union general assembly [Ciolfi et al. 2016a]. some results of the symbiosis study (chapter 4) have been presented at the first Italian stable isotopes IRMS meeting [Ciolfi et al. 2016b].

⁹The results are published on Food Chemistry [Chiocchini et al. 2016].

1 State of the Art in Space and Time Modeling

Ecological modeling is a multifold subject. It includes a variety of techniques borrowed from countless fields of research, each one characterised by its own rules. Complex systems show many sources of variability at any scale of reference; often we have nested cycles which vary both in space and in time, cyclically or not. Such a complex environment calls for compromise: on the one hand, the constraints of a clean statistical procedure are often too strict for ecological modeling so that we must somehow loose them, on the other hand one has to aim at the highest mathematical rigour available to strengthen his/her hypotesis.

The widespread availability of Geographical Information Systems (GIS) brought many once highly specialized tools to a broader users pool. GIS systems integrate remote sensed data with the users' own measurements; also, a GIS is a safe environment for data storage. Statistical packages like R offer countless tools for almost any calculation and some GIS capabilities, too.

However, there is a lack of simple tools for mixing freely spatial and temporal variability, which is peculiar to many ecological dataset, due to the nature of the samples' collection.

1.1 Spatial Modeling

Spatial modeling is a wide subject. Basically, it allows the reconstruction of spatial patterns from the input dataset. Thematic maps can be build as surfaces from many ecological variables [Fortin and Dale 2011], supposing there is uniformity in time, i.e. when one can treat the data as if they had been collected at the same time [Wagner and Fortin 2005].

Since space interpolation is a consolidated subject, there is widespread consensus on the actual techniques, like Inverse Distance Weighting or Kriging [Cressie 1990]. Unfortunately, sometimes the peculiar nature of ecological datasets places them out of the realm of applicability of most interpolators, but the common practice suggests to proceed with the interpolation,¹⁰ using extreme caution when interpreting the results [Rossi et al. 1992]. Any *variable estimate* should be accompanied by its *residual estimate* that testifies how trustworthy the output is: a sort of second-order or meta- statistics [Cressie 1990]. Estimating a thematic layer without its residuals is like quoting measurements without errors, but surprisingly many authors are loose on this subject [Goodchild et al 1992].

Pedology, which is the most “geological” subject in ecological modeling, is perhaps the discipline where most of the methodological efforts have been done [Krasilnikov et al. 2008].

¹⁰The best practices suggest to use an interpolation algorithm which is capable to estimate the error or the confidence interval of the output model. Kriging is often the estimator of choice also because it computes automatically this error estimate [Oliver and Webster 2015].

Spatial variability exists at all the possible geographical scales; ecological studies are often limited to a relatively small area (ecosystem level), but broader areas are often covered, up to the whole planet. Global models introduce the further complication of dealing with curved surfaces, the earth as a whole needs to be represented via angular coordinates [Goovaerts 1997], but it is only a question of computational power, which nowadays is largely overcome also by standard desktop computers.

The recent availability of (open or proprietary) GIS packages brought ecological geostatistical modeling into a GIS framework. Mapping and remote sensing are an important part of many researches, so working within a GIS context is always a good option. The *open source* community offers a lot of options, including the comprehensive Qgis environment¹¹, which is the reference standard [QGIS]. A widespread commercial alternative is ESRI's ArcGIS.¹²

Within the boundaries of *open source*, the R statistical package¹³ provides almost every useful function [R]. Specialized packages exist for many ecological issues [Angerosa et al. 1999]. There exists a sort of *numerical ecology* based on R [Borcard et al. 2011], which also integrates a basic set of GIS and graphical functions. Added benefits of conducting the spatial modeling within R are the protection of the dataset and the straightforward access to a countless set of trusted statistical tools. Working in R ensures the best matching with the current so-called *geomathematical* algorithms [Agterberg 2014, Bivand et al. 2008]. Of course, also non spatially-related ecological datasets can be conveniently treated in R [Zuur et al. 2007]. Many modeling techniques from the broader subject of geosciences can be integrated as well [Sarma 2009].

The geostatistical toolbox consists fundamentally in two families of estimators: deterministic interpolators such as the IDW (Inverse Distance Weighted interpolation) or Bayesian as the Kriging inference algorithm.¹⁴ The Bayesian approach is gaining popularity in ecological modeling [Diggle and Ribeiro 2007]. These modeling techniques can be applied more or less strictly according to the actual samples involved in the research [Renard et al. 2010].

¹¹It is appropriate to say *environment* instead of, simply, *software*, since Qgis is in fact a framework that includes many modules written in different programming languages. Qgis integrates a collection of GIS-related software from nearly all open source projects.

¹²A non-open source software, in principle, uses unknown calculation procedures which within a scientific workflow should be avoided. Anyway ArcGIS is *de facto* an established standard in forest ecology as well as other environmental sciences fields. The software documentation provides some general notes about the actual algorithms employed. ArcGIS is a bit friendlier than its open counterparts, but it is limited to Microsoft Windows systems. It also produces the most pleasing maps, aesthetically speaking. See [ARCGIS].

¹³The R learning curve is notoriously awkward, but the benefits of a solid statistical package are countless. Those new to R should consider the RStudio graphical user interface [RSTUDIO].

¹⁴The ubiquitous Kriging consists in fact of more than one estimator. The choice depends on the distribution (variogram) of the sampled data. Not all the datasets are suitable for Kriging, though: some requirements of stability must be satisfied.

1.2 Temporal Modeling

As a matter of principle, in ecology there is not a single phenomenon without an evolution over time. We speak of time variability and consequently of time modeling when time is the prevailing driver of change, i.e. when the variability with respect to time is much greater than that over space. Actual time series are replicas of the same measurement, but in real-life experiments the (quasi) replicas are inevitably affected by countless noise factors, so the subtleties of time series analysis should be applied with some care [Cryer and Chan 2008]. In the last years the techniques of fuzzy logic and pattern recognition have entered the realm of time series [Pedrycz and Chen 2013] but no applications to ecological modeling has been done as yet, while the more conventional tools of Fourier analysis have been applied successfully [Bence 1995, Turchin and Taylor 1992].

Time modeling is particularly useful when seasonality is at work: this is precisely the case in many ecological studies [Box et al. 2016, Jassby and Powell 1990]. In any case, it is very difficult to compare the measurements taken at different places after the machinery of time series analysis has been applied.

The modern instruments often produce a continuous output; this is the case, for example, of meteorological stations, perhaps remotely controlled. This is good, of course, but everything comes at a price: measurements at different sites need to be uniformed before treatment and, above all, measurements are redundant and it is not always clear how to extract meaningful averages from them [Liao 2005].

A few measurements repeated over time do not make a time series, however. There is a lack of techniques to deal with few, sparse, time samples (Fourier transforms and algorithms of the like require a large amount of data). In these cases finding a temporal pattern can be extremely difficult, if meaningful at all.

Other than spreadsheet handicraft, classical time series analysis [Hamilton 1994] can be applied in R, which is again the ideal playground for ecological modeling [Stevens 2009]. A limiting factor in the treatment of time series is the number of records, which is often beyond the reach of ordinary spreadsheet programs¹⁵ so that the storage in a full-fledged database should be taken into account.¹⁶

¹⁵The main concern with spreadsheets (Excel & co.) is the scarce *protection* offered to the data. On the other hand, it is advisable to perform a tentative, fast, analysis with a spreadsheet for a first glance at the general behaviour of the dataset. Some useful techniques are illustrated in [Legendre and Legendre 2012].

¹⁶Switching to a proper Relational Database Management System (RDBMS) like MySQL or Oracle for data storage comes not for free, in terms of computational effort [MYSQL, ORACLE], but it adds many benefits in terms of data protection and it opens up the access to a number of advanced instruments of analysis, the so-called *data mining* machinery [Nisbet et al. 2009].

1.3 Spatiotemporal Modeling

Strictly speaking there is not pure spatial nor temporal modeling in ecological systems, due to their complexity. Every evolutionary phenomenon brings spatial as well as temporal evolutionary patterns: it is a matter of choice of the researcher to prefer one driver of change with respect to the other. Sometimes, however, spatial and temporal variability has to be accounted for on equal grounds and, unfortunately, there is a lack of off-the-shelf instruments for doing so in the standard numerical ecology workflow (pretending such a *workflow* exists at all).

Physics has encountered the same problems long ago, since the beginning of the 1900s and, as of now, there is a number of consolidated, mathematically unexceptionable, techniques to deal with *spacetime* variability [Szeckeres 2006]. The geometrical approach is centred about the concept of spacetime that, since the first works on Einstein's Relativity¹⁷ has become the cornerstone of modern physics.

The field of statistical physics is a source of techniques which in principle can be adapted to the needs of ecological modeling: ecological systems are, at the vey end, extremely complex open thermodynamical systems [Sethna 2006]¹⁸. Though this is enough to justify on mathematical grounds the use of spacetime techniques in ecology, it is not of any practical benefit.

Ecological systems are centred on the complexity of the relationships, more than the subtleties of mathematics, this brings pros and cons into the play: on the one hand we can be a bit loose when checking all the statistical constraints, on the other hand we cannot expect the same bombproof predictive power from our models. As the pioneering works of Christakos have shown there are not actual methodological obstacles in placing ecological modeling into a spacetime framework [Christakos 2000], even in a Geographical Information System context [Christakos et al. 2002].¹⁹

The first algorithm to be converted from a space-only to a space-time version was Kriging [Bogaert 1996]. This is no surprise since its known performances and its capability to estimate errors, other than values. The interesting review of Kyriakidis and Journel also favours Kriging over other geostatistical techniques, following the general Bayesian inference fashion that pervades almost every research work [Kyriakidis and Journel 1999]. This is not to say, of course, That Kriging is not one of the best techniques, but it is sometimes applied blindly, avoiding

¹⁷Physics employs a variety of spacetimes: classical, relativistic, quantum etc., according to the geometric and energetic scales involved [Frankel 2012].

¹⁸Most of statistical physics is devoted to quantum systems, which are of no interest in ecology, that deals with classical systems. Techniques exist, however, for the treatment of classical systems inherited from their quantum counterpart [Schlosshauer 2008].

¹⁹Christakos does not suggest practical algorithms. In the first 2000s GIS technology was not quite widespread and the computing power availability was comparatively modest.

other (maybe simpler) techniques which could do a good job too.

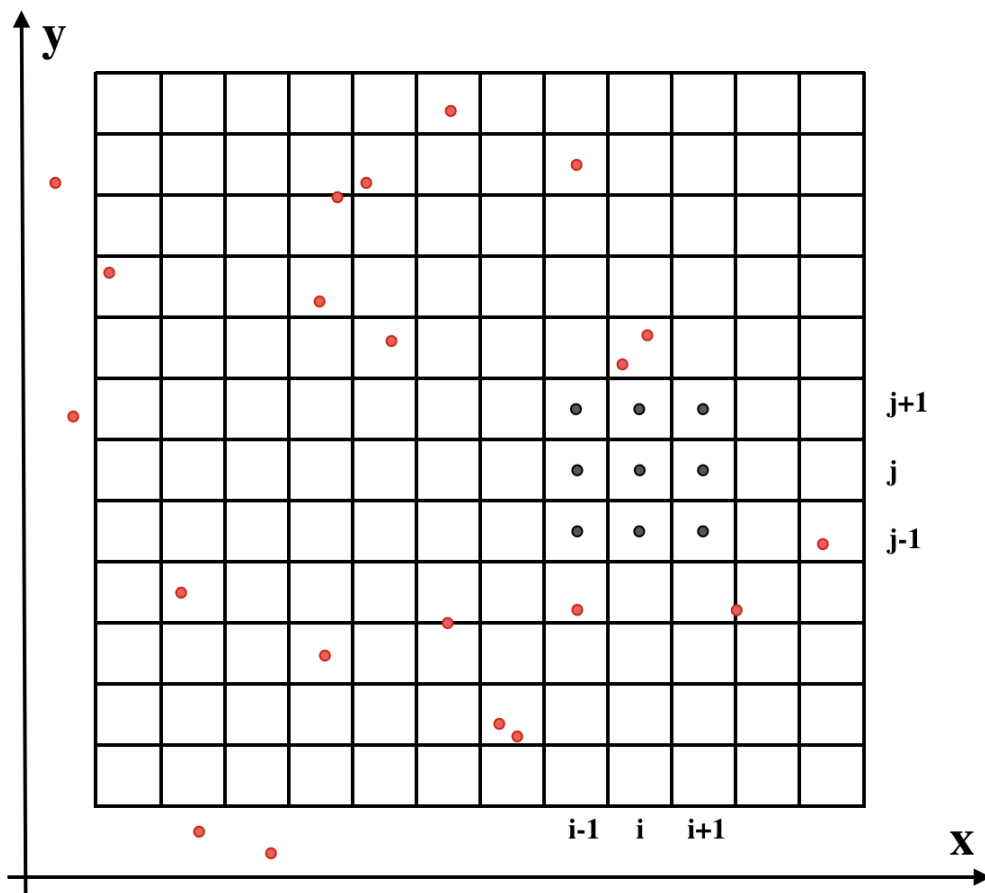
The updated work of Cressie and Wikle also illustrates Kriging and other estimators “upgraded” to the spacetime version. In all cases the basic step is the transformation of time into a new space dimension [Cressie and Wikle 2011]. Operatively, the operation involved is a multiplication of time for a constant factor. This factor has the dimensions of a velocity and sometimes it is related to an actual velocity of a transport phenomenon; unfortunately, this analogy should not be pursued literally in most cases. A detailed mathematical discussion of the role of the conversion factor can be found in Appendix A.

The Timescape algorithm follows the general trend of transforming time into space, adding a causal structure, tailored to the needs of the researchers, according to the phenomenon they are investigating. A causal structure [Cressie and Wikle 2011] is a deformation of the space (better, the spacetime) which accommodates the relationships among the measured quantities. Nowadays we face almost an excess of computing power and software resources, compared to the needs and consistency of most ecological studies datasets. There is no reason for avoiding the simultaneous variability in time and space if the research needs both of them.

2 The Timescape Algorithm

Spatial statistics has a few controversial issues at its very foundation. On the one hand, statistically speaking, data autocorrelation is bad; on the other hand, it is a founding requirement that closer sites should have similar values of space-related quantities, and this is a form of autocorrelation. Adding the further complication of a third dimension does not resolve these issues but, at least, it does not make the things worst.

The simplest geostatistical algorithm, the IDW (Inverse Distance Weighting) assigns a sites' value of a certain quantity through a weighted average of some known values (the red dots in the figure below, which can be, or not, arranged according to a regular lattice).



The model is cell-based. The cells (or pixels) are arranged according to a regular lattice; the model's sites are the centres of such cells (the black dots of the figure, only nine dots are shown not to clutter the figure too much). Each model site is indexed by an (i, j) integer coordinates pair. Different model have different pixel sizes, i.e. different spatial resolutions, and the model size is MN , where i ranges from 1 to M and j ranges from 1 to N .²⁰ Many interpolation strategies exist in literature [Cressie 1990] according to the dataset nature and to the kind of modelled phenomenon.

²⁰This is a general characteristic of any finite model, not only of IDW.

An important fact to be stressed is that the *complexity* of the model scales as N^2 or as r^{-2} , where N is the typical cell count per axis and r is the spatial resolution or *pixel size*.²¹ The computation time varies according to the number of cells, so it goes as N^2 as well.

Time series, on the other hand, are one-dimensional problems, computationally speaking. So we use to say that their complexity scales as N . Now putting a time series on top of each space cell causes an enormous growth in complexity, and as a consequence a comparable growth of evaluation times. A space-time model scales as $N^2 \times N = N^3$. Just to give an example, if the model resolution is typically about $N = 1000$ per space (and time) direction, as a respectful image is, the value of N^3 is one billion (compare to the standard digital picture sizes of a few million pixels). The bulkiness of such an amount of data was simply too much for an ordinary desktop computer, up to a few years ago.

Last but not least, the manipulation of time series of data follows its own rules, which do not always coexist with space statistics without trouble [Cressie and Wikle 2011, Hamilton 1994].

Many ecological datasets are collected within an enclosed area (spatial distribution) and over a temporal interval (time distribution). These sources of variability are both present at the same time and in most cases they are interlinked. The most common solution, especially if the data collection follows a periodic (say seasonal) time pattern, consists in evaluation separate space models at different times, as if we had distinct datasets.

Sometimes this is an acceptable solution, sometimes not. If the data collection has taken place while the sampled values were changing, or if the collection times are uneven (not periodically) spaced, we cannot arbitrarily assign one sample to a certain plane. Furthermore, at any time we must have a conspicuous number of samples, for the interpolation to be done.

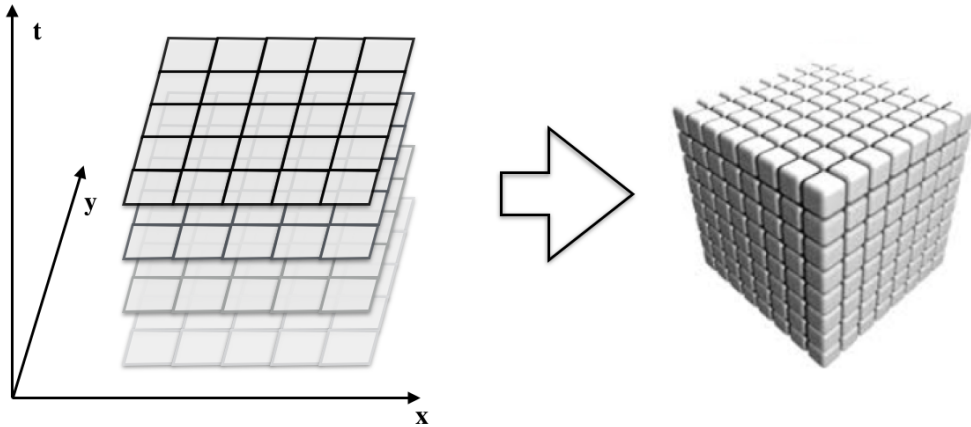
A yet simpler solution exists: neglecting time altogether. It works only with space-related *stationary* patterns, which is almost never the case.

The Timescape proposal allows one to use the ordinary machinery of geostatistics (IDW as said, or Kriging or any other space-only interpolation algorithm) with some pre-treatment: the so-called spatialisation of time. This is nothing new in the fields of field theory and statistical mechanics, from which the idea is borrowed, but the kind of input dataset is completely different. Data from forest ecology, by the very nature of the ecosystems involved, are affected by countless

²¹As a matter of principle, pixels need not to be squares. But dealing just with the order of magnitude, it is customary to use the same figures for all the space directions.

sources of noise. We cannot expect the precision of the lab-based, controlled environment experiments but, on the other hand, the bonds of statistical rigour can be somehow loosened.

The Timescape idea in a nutshell consists in *spatialisation plus causality*. Spatialisation of time is what allows us to use the machinery of spatial statistics, but time cannot be simply regarded as a third direction of space, it is a forward-only direction, so a *causal structure* has to be superimposed.



A Timescape model is not simply a collection of parallel planes. It is a collection of *voxels* (volume elements): it possesses a full three-dimensional structure. The peculiar nature of the vertical temporal dimension is encoded in the causality constraints that are imposed during the evaluation of the model. It is clear from the figure above how performing some surgery on the finished model we can slice two-dimensional spatial models and we can also dig one-dimensional cores as actual time series approximations (see section 3 for details about the kind of data that can be extracted from the local and global versions of the software).

2.1 Evaluation of Distances

The spatialisation of time is very easy. It is just a matter of multiplying the time coordinate times a constant c having the dimensions of a velocity:

$$(t, x, y) \rightarrow (ct, x, y) \quad [c] = \text{LT}^{-1}$$

It is worth mentioning that c is related to *how fast* the values of the measured quantity can change, it is not necessarily related to the actual flow of anything physical: it is the *information* which flows from voxel to voxel. Of course, if there were a transport phenomenon occurring, c could be its velocity. The value of c can be a constant or, more generally, it can be related to the position and time (see Appendix A for details) in the following it is considered constant.

All geostatistical algorithms are based on the notion of distance. There is not the distance, however, but a whole class of suitable functions²² which can be chosen according to the user's needs. The simplest distance function is the ordinary Euclidean distance, which is simply the multidimensional version of the Pythagorean theorem:

$$d(A, B) = \sqrt{c^2(t_A - t_B)^2 + (x_A - x_B)^2 + (y_A - y_B)^2} \quad (1)$$

that gives the measure of the distance between points A and B , including the temporal component. Equation (1) can be used only for projected coordinates (UTM, Lambert, etc), it is the distance of choice of the **TimescapeLocal** software (see section 3.1); x and y are the horizontal and vertical coordinates.

This expression of the distance is not valid in a curved space, as is the case of the Earth surface. In this case we must use an estimate of the ground distance as the measure of the geodesic arc connecting the A and B points on the surface of the Earth. The **TimescapeGlobal** software (see section 3.2) estimates such a distance as the arc length of the maximum circle connecting A and B :

$$d(A, B) = \sqrt{c^2(t_A - t_B)^2 + R^2 \arccos^2(\sin \varphi_A \sin \varphi_B + \cos \varphi_A \cos \varphi_B \cos(\lambda_A - \lambda_B))} \quad (2)$$

where λ is the longitude, φ the latitude and R is the radius of the Earth, approximated as a sphere.²³ Equations (1) and (2) are complementary in that they are suited to different space scale problems. The first is more useful in the detailed (large scale) studies typical of forest ecology, while the latter is best suited to small scale, say worldwide, cases. It is of course possible to evaluate the distances within a few hectares as geodesic arcs, but it is a complete waste of time in terms of computation complexity.

A Timescape model is a discrete collection of coordinates (t, x, y) of the centres of a lattice of voxels. For any element of the model it is possible to evaluate the distance d from any point of the source dataset. We then use these distances as weight estimators for calculating the values of the model elements.

It is possible to use the ordinary spatial statistics techniques in three dimensions [Cressie 1990,

²²A *distance* $d(\cdot)$ on a set X is a function $d : X \times X \rightarrow \mathbb{R}$ which is nonnegative $d(x, y) \geq 0$, symmetric $d(x, y) = d(y, x)$, coincident $d(x, y) = 0$ iff $x = y$ and subadditive $d(x, y) + d(y, z) \geq d(x, z)$. See Appendix A.

²³Better shape approximations for the Earth, for example as an oblate spheroid, would increase enormously the complexity of the calculations, without a real benefit on a continent- or world-scaled problem. See Appendix A for details on the possible metric functions on curved spaces.

R], using the time as a correction factor only. It is sometimes useful (better than nothing, one would say) but it does not catch the peculiarity causality as it develops over time. We must find a way to introduce causality in our models.

2.2 Causal Constraints

The simplest interpretation of causality says simply that previous events can be the cause of subsequent ones, but not vice versa. This has an immediate translation into a *causal constraint*: a voxel value is calculated only using its foregoing sample points, neglecting the others.

This basic form of causality can be immediately implemented as a three-dimensional spatial statistical technique: it is ordinary interpolation (IDW, Kriging or whatever) with a flexible input dataset, just picking the appropriate points for any model voxel. The **TimescapeLocal** and **TimescapeGlobal** implementations of the Timescape algorithm are able to run models with this primitive form of causality. It is also possible to adapt the R package algorithms [R] for such a job, simply adding to the samples dataset a flag for each model element for picking the right source points.

We can also introduce a more complex *causal structure*. The leading concept is that of an *expanding zone of influence*, i.e. a point of our spacetime can influence other point in its future (but not in its past), provided these points are not too far; how far depends on time, it is like a process of diffusion. Think for example of a drop of ink on absorbent paper: the blackness of the paper grows over time at a definite speed.

In Timescape, we model the spreading of the measured values as if it was a diffusion process. The c parameter, as said beforehand, is the rate of diffusion of *information*, pretending that the source points dataset is diffusing values into the model. Mathematically, it is a diffusion process, also if there is nothing physical actually flowing anywhere.

We have to put limits on the influence of the source points. As pictured up to now, the role of the c parameter is simply that of a conversion factor, there is nothing preventing an instantaneous propagation from point to point of our model. This is far from any actual process, where it takes some time for a perturbation to be felt far away.

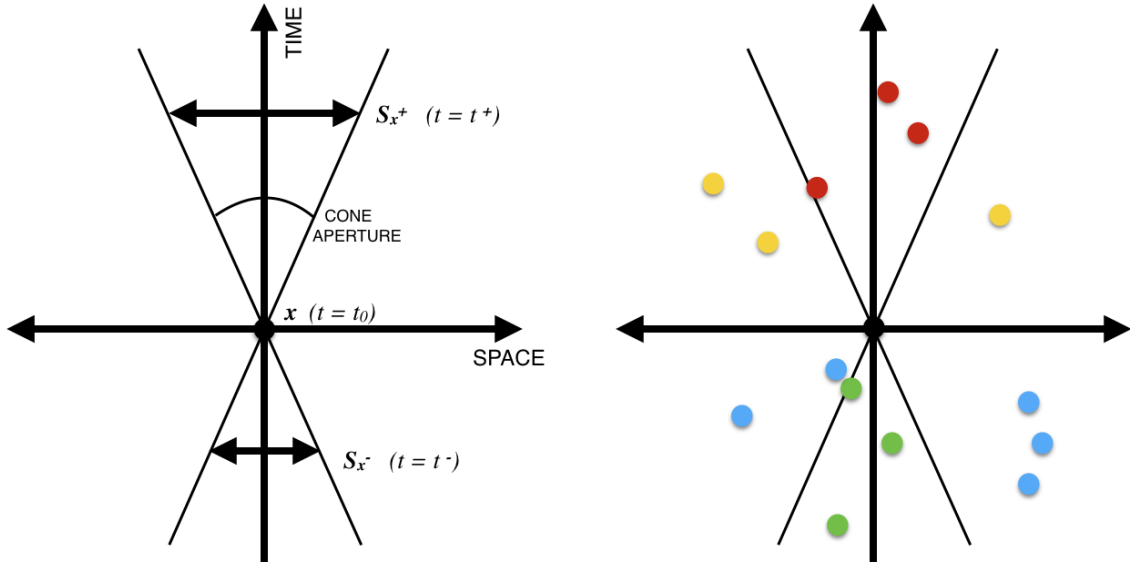
The notion of spacetime in physics dates back to the beginning of the 20th century with the Minkowski formulation of the Special Relativity Theory. Since then, the spacetime approach has been proven efficient in field theory and statistical physics as the ideal arena for the description

of the most diverse phenomena. Spacetime is the stage of changing patterns in many other disciplines and we believe that it could be useful in forest ecology modeling as well, despite some mathematical subtleties.²⁴

Timescape borrows from the spacetime of Physics the notion of causality.²⁵ In relativistic Physics the notion of *light cone*, strictly linked to causality, emerges from prime principles; this is not the case in the spacetime described above, with the ordinary Euclidean distance. In our case a causal structure has to be “enforced by hand”, imposing a constraint of maximum possible influence. We define the *causal ratio* r_{AB} as the adimensional ratio between the ground distance and the time distance of the points A and B , i.e.

$$r_{AB} = \frac{\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}}{c |t_A - t_B|} \quad (3)$$

this quantity measures how far (in space) A and B are with respect their separation in time. The factor c is needed to keep the quantities comparable. The figure below shows the space dimension(s) horizontally and the forward-only time dimension vertically.

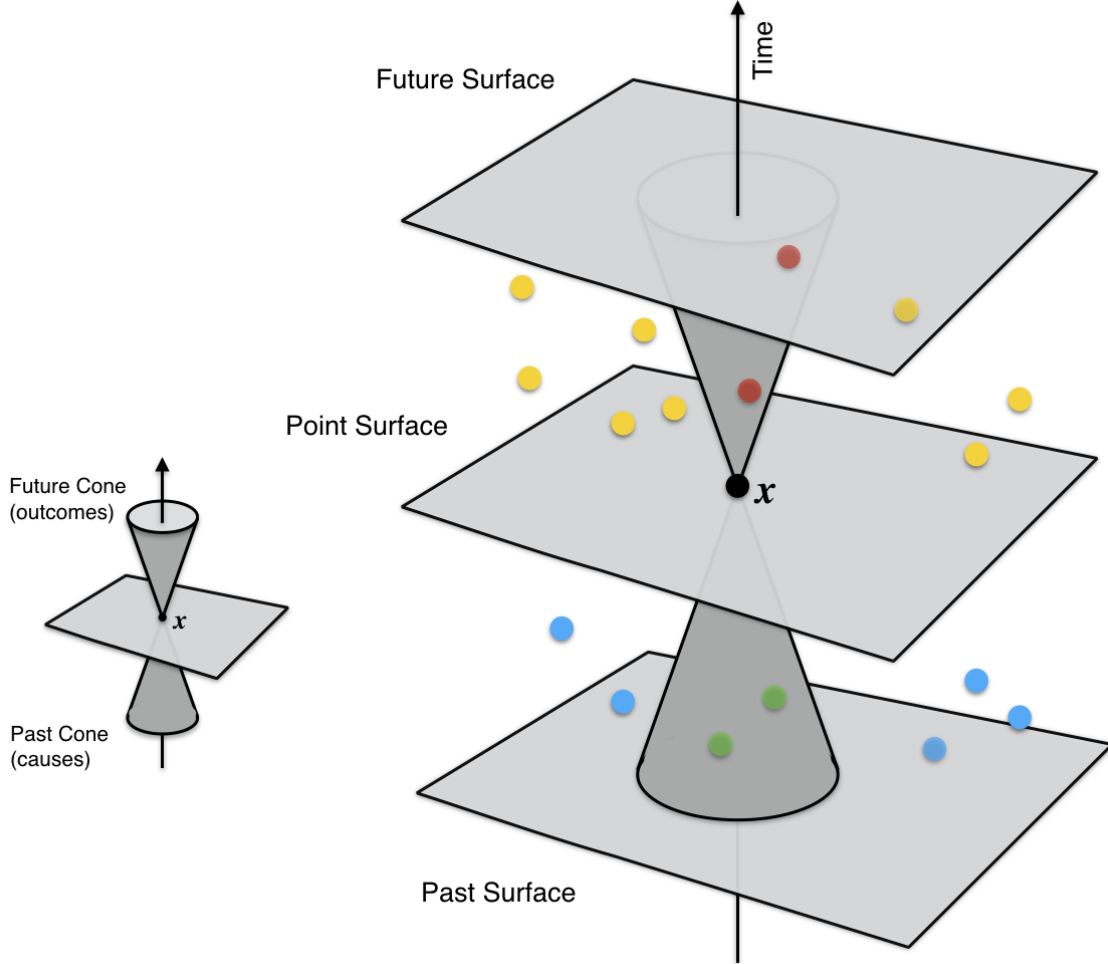


Enforcing a causal constraint means defining a certain cone of influence on such abstract space. A *forward causal cone* contains all the possible outcomes of a source point x (red dots), while the yellow dots indicate unreachable events, i.e. they cannot have been influenced by x . In the same way, green dots belong to the *backwards causal cone* so they are possible causes for

²⁴The Minkowskian spacetime is not what is going to be used in Timescape. Nor Timescape has anything to do with Relativity, even if there are -deliberately- some common mathematical techniques involved.

²⁵Causality emerges naturally in a relativistic framework as a consequence of a maximum allowed velocity, the speed of light c , thus dividing the spacetime of a point in accessible (so called *timelike*) areas and unreachable (*spacelike*) ones, even for the light itself.

x , while the blue dots are not. The aperture of the cone can be adjusted according to the very nature of the modelled phenomenon, so for any point x of the model at a time, say, t_0 , we have a moving forward surface S_x^+ which scans the future of x at any time $t^+ > t_0$ and a backwards surface S_x^- which scans the past of x at any time $t^- < t_0$.



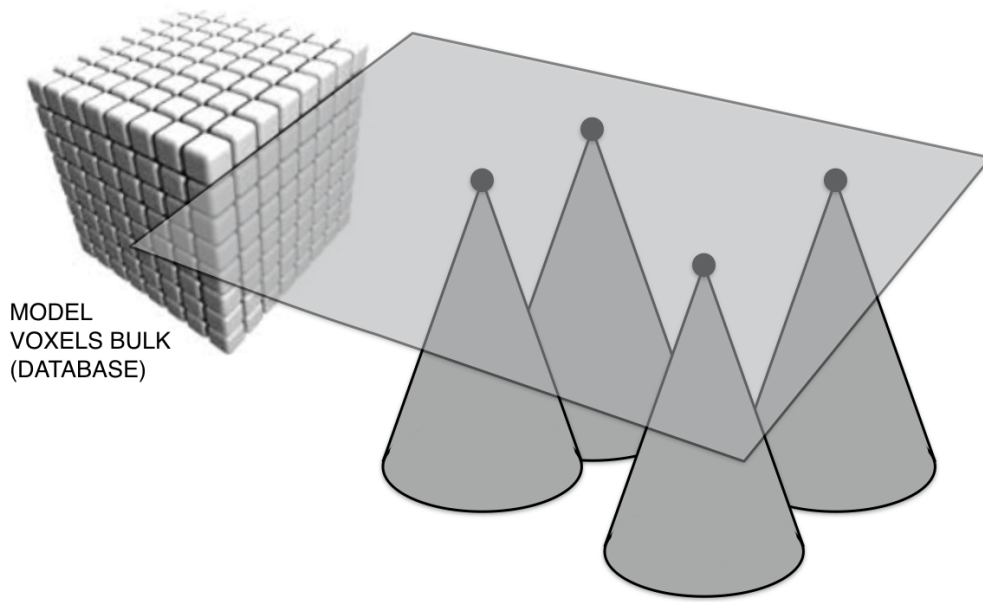
Inflating the two-dimensional construction to a full three-dimensional representation, as in the figure above, we can see that every point x of the model possesses a causal structure. In particular, following the same colour coding, the green dots are the set of events (a subset of our observations) which could have caused x , so we must estimate the our value at x using only its green dots, i.e. the elements falling into x 's backwards causal cone.

How wide should the cone be? it is controlled by the Timescape parameter k , which is the maximum acceptable value r_{AB} for two points A and B to be causally connected (eq. 3). Operatively we have just to check $r_{AB} \leq k$; which point is the cause and which one is the outcome depends on the values of time:

$$\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \leq k c |t_A - t_B| \quad (4)$$

Large values of k are related to loose constraints, i.e. a fast diffusion is allowed.²⁶ on the other hand, small values of k are related to very strict causal constraints, that is low diffusion rates. The value of k is assumed constant here for the sake of simplicity, but it can be a function of the position in the spacetime and also of ancillary variables (external conditions). The mathematical appendix (A) describes in detail all the subtleties about k .²⁷

The particular flavour of a Timescape model is given by the combined values of c and k . c tells us how much we must correct our readings with respect to time differences, while k says how strict we are about causality.²⁸



2.3 Model Evaluation

The Timescape model is built as follows: an empty model is inserted into the database then, one *sheet* (constant time section) at a time, all the voxels are evaluated. Each sheet's voxel has its own causal cone which contains only a subset of all the samples from which we are interpolating. If the cone is empty, no value can be attributed to the voxel. After the interpolation of the voxels' values, these are re-inserted into the database, and the calculation proceeds with the next sheet, till the last.

²⁶An infinite value of k is acceptable as well, meaning that the causal cones fill all the space.

²⁷The **TimescapeLocal** software (section 3.1) allows only constant k values, including 0 and ∞ , while in **TimescapeLocal** (section 3.2) k is a user-defined function, which is by default a constant.

²⁸Technically speaking, it is a Wick-rotated Euclidean version of Minkowskian spacetime but, unlike the latter, it requires two separate parameters to establish the causal relationships among points. This construction resembles the *light cones* of Special Relativity: as said, this is not by chance, but the realms of possible applications of the two constructions are completely different. Timescape is thought for ecological issues. The interested reader can consult any manual about Special Relativity; an easy, accessible one is [Woodhouse 2003]. The technical term for a point of spacetime is *event*, as in Appendix A, but we will not follow this convention here due to possible confusions with the layman meaning of the word *event*.

If the backwards causal cone of the voxel located at a certain point x contains at least one source point, the evaluation can take place as follows:

- Distance assessment: for all the source points s_k evaluate the corresponding distances d_k from x , thus obtaining the couples (s_k, d_k) . Each source point s_k possesses a value v_k .
- Pruning: order the couples by increasing distance and retain only the first N of them. It is possible to skip this phase and use all the couples. Usually just the few nearest s_k give a significant contribution.
- Spatial statistics: use any established geostatistical algorithm to estimate the value at x . For example, using a simple IDW, where the weights $w_k = 1/d_k$, the value v at x is:

$$v = \frac{\sum_k w_k v_k}{\sum_k w_k}$$

The last step ensures that Timescape is at least as good as the spatial interpolator adopted, inheriting from it the proof of convergence (and the interpolator's defects too, of course). As of now, there are lots of options for the spatial statistics step, other than the plain IDW: **TimescapeLocal** offers a variety of methods, ranging from IDW to Kriging, incorporating also harmonic (periodic) corrections,²⁹ while **TimescapeGlobal** is based on a single flexible definition of weights, but does not offer Kriging. See chapter 3 for details.

In fact, the Timescape Algorithm does not act on interpolation itself, but on the structure of the samples dataset with a suitable distance definition [Cressie 1990]. It is like reshuffling the dataset for each interpolated voxel, so that each voxel is a little model on its own. This procedure is fair as long as the distances defined are true distances.³⁰

2.3.1 Model Tuning

There are many options for Timescape tuning, some were already mentioned before, about distances and causality. The distinctive character of any Timescape model is given by c and k , since these model the space-time relationship according to the user's needs, but how to choose these parameters if there is no clue about their values? As a rule of thumb, as an initial guess c should be chosen in such a way that the model bulk is roughly a cube, i.e. $c \Delta T \sim \Delta X$ or $c = \Delta X / \Delta T$, where ΔX and ΔT are the space and time intervals of the model; if they are

²⁹It is possible to plug one's own interpolators, but it requires the coding of a Java class per each new interpolator, extending a suitable abstract class.

³⁰See the footnote on page 17 for the mathematical definition of distance.

not comparable, probably space or time variability is the prevailing aspect by far, so one can use ordinary spatial statistics or time series analysis techniques. A first guess for k is harder to motivate, one can simply try $k = 1$ and move towards ∞ or 0 if there are too few or too many points falling inside the causal cones.

If one knows about a transport phenomenon which is occurring about the variable under scrutiny in the complex ecological system, c can be chosen as the appropriate velocity (in units of length per units of time³¹ in **TimescapeLocal** or in degrees per unit of time in **TimescapeGlobal**), in this case a good guess for k should be 2 or 3; $k = 1$ can be chosen if one knows for certain that c is an insuperable threshold.³²

The fine tuning of a model is achieved through a set of other parameters, which include:

- The *neighbourhood consistency* N : it is the maximum number of source points to be considered for the statistical interpolation.
- The *maximum distance* D , if we assume that the source points more than D apart cannot have any influence on the estimate.
- The *metric* employed: the neighbourhood of a point can be considered a circle, a square or a diamond.³³ It does not affect the general behaviour of a model but can change the values.

It is also possible to interpolate the values from ancillary variables instead of the actually measured values, or to corroborate the estimates using both measured and ancillary values. Users need to define their own interpolation functions using a simple syntax.³⁴ A note of caution is in order here, however, since the constraints of a well-behaved distance (footnote on page 17) have to be satisfied, or there is no guarantee that the spatial statistics methods employed can converge.

The number of voxels which constitute the Timescape model can be defined, ranging from a few thousands for little downscaled models to billions, depending on storage capacity.

³¹There are no predefined units in **TimescapeLocal**, space can be given in metres, as a general rule, for this is the unit of UTM and Lamber Conical projections. **TimescapeGlobal** uses decimal degrees in place of proper length units, but internally the lengths are calculated in kilometres, so c has to be given in kilometres per unit of time units.

³²This is the case in Special Relativity, where c is the speed of light and no other parameter is needed to define causality.

³³These are said to be *equivalent metrics*.

³⁴As of now, functions must be implemented in Javascript [ECMA], which is fast and easy to use also for non-expert users. Most algebraic functions can be implemented as-is in textbook notation.

2.4 Model Storage and Exploration

Admittedly, a real drawback of Timescape is the need of a lot of room for model storage. This calls for an external database. An external storage, however has its advantages too. The model can be queried in standard SQL³⁵ (Structured Query Language) for any need that users can have. To keep the bulk to a minimum the actual coordinates of the voxel are not stored into voxels table, which contains only their references, i.e a voxel is represented as a record

$$\text{voxel} = [k, i, j; \text{value}]$$

where (i, j) label the space site (the “horizontal” coordinates) and k labels the time sheet (the “vertical” coordinate). It is up to the user to get back to the real-world coordinates and time.

A voxel record is so composed of three integer numbers, the coordinates, and a real number,³⁶ the value. The indices (i, j, k) range from zero for minimum x /longitude, minimum y /latitude and time, and their maxima N_x , N_y and N_t , corresponding to the maximum value of the coordinates. The total number of voxels is $N_x N_y N_t$, which corresponds roughly to a size of $20 N_x N_y N_t$ bytes. Just to give an idea, a small $100 \times 100 \times 100$ model needs about 20 MB and a $1000 \times 1000 \times 1000$ one requires 20 GB of database space. This is the main reason for choosing an external database for models’ storage.

The Timescape published software versions offer a variety of pre-packaged querying tools.³⁷ These include statistical analysis tools and allow the export of different subsets of the model.

Timescape can be thought of as a “detour” from the standard users’ flow, combining GIS and statistical analysis. To this end, it is possible to export the time sheets as GRID files, a common GIS raster standard (they are human-readable ascii files), and cores dug at at given site, which are in fact time series of modelled values.

It is also possible to export data in form of *.csv* files (comma-separated formatted ascii files). This is the standard input of any serious statistical package; *.csv* files can be imported in most spreadsheets, too, but the quantity of records discourages such approach for the bigger models.

The Local and Global versions of Timescape offer different choices for exporting the data, according to the peculiar needs of the scales of reference.

³⁵The standard distribution of **TimescapeLocal** and **TimescapeGlobal** is based on a MySQL database [MySQL]. The database connection, however, is mediated by a Hibernate framework [HIBERNATE], so users can adopt any Hibernate-compatible database, like Oracle [ORACLE]. See Appendix B for details.

³⁶Technically, the value is a so-called *double* number.

³⁷See section 3.1 and 3.2 for details.

3 The Software Packages

The Timescape project has been split into two branches, **TimescapeLocal** and **TimescapeGlobal**, according to the problem scale and to the space coordinates employed. The local version works with projected coordinates and it is best suited to detailed scale studies. It is perhaps the most useful in forestry and ecological studies.

The global version uses geographical coordinate (latitude and longitude). This implies a lot more calculations for the geometrical part, since the distances have to be evaluated as lengths of geodesic arcs on the surface of the Earth. This version is best suited, as the name suggests, for global scaled studies, where the whole Earth, or at least a continent-sized area, is examined.

In both cases data should be stored into an external database,³⁸ one database per dataset. This limitation is due to the bulkiness of the models, which is in the order of GigaBytes per model. It is also a choice of cleanliness in the data storage model, which keeps to a minimum the cross-references between the tables.

The input dataset has to be a collection of values for which one knows the value that should be used for the interpolation, along with the horizontal and vertical coordinates, and the time of observation. Other (ancillary) data values can be associated to each observation, these can be used to improve the reliability of the calculated values.

Both softwares share a common workflow: users should create a fresh model, define its parameters, run it and then they can explore the finished model and export a variety of information from it. This procedure can be inserted smoothly in the users' ordinary geostatistical workflow, since the input and output procedures are kept as standard as possible, using ascii files and GIS-standard layers as output.

On each database can be stored as many models as the user needs, the limit is only given by the actual database capacity. This is one of the reasons, if not the most important benefit, for using an external database for storage, although it adds some extra work for the installation of the software. To ensure the maximum possible hardware/software compatibility the programs are in Java language [JAVA]. This choice is due to the availability of Java Virtual Machines for any desktop operative system and to the Java-database(s) reputed dialoguing ability. The distribution includes a pre-packaged *jar* (java archive) executable file.

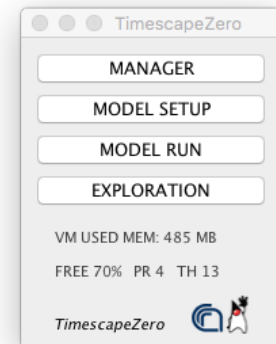
³⁸ *External* does not mean necessarily a dedicated server, it can be run on the same machine of the application. It is just *external to the application*.

3.1 TimescapeLocal

The local *flavour* of Timescape is **TimescapeLocal**. It is the first branch of the project, which operates on projected coordinates. The software does not take care of the actual projections used, as long as they are compatible with the distance functions provided (Euclidean or equivalent). UTM (Universal Transverse Mercator) or Lamber conic projections work well, as many other less used coordinate systems. The output coordinates values follow exactly the input ones, for the sake of compatibility with other georeferenced data that the user will have to work with in his/her project.

The main window is a sequence of buttons, logically ordered following the create-define-evaluate-explore model workflow sketched beforehand. Each button pops-up a dialog window for the user to operate on his/her models. At the start, if the database is empty, the program asks for a formatted dataset to parse and store.³⁹

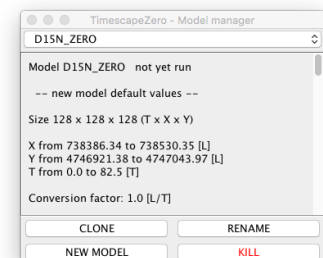
Each button corresponds to a logical function, detailed below. Other than the **MANAGER** section, there is a **SETUP** section for defining the model parameters (all fresh models are created the same, with default parameters); a model is then **RUN** in a separate window and, upon completion, it can be examined or explored. From the **EXPLORATION** panel as well the user can export various subsets of the model.



3.1.1 The MANAGER Panel

This panel manages the model-level actions which are:

- model creation
- model renaming
- model cloning
- model destruction



The meaning of each operation is self-explanatory. The **CLONE** function is particularly useful to define a collection of models with

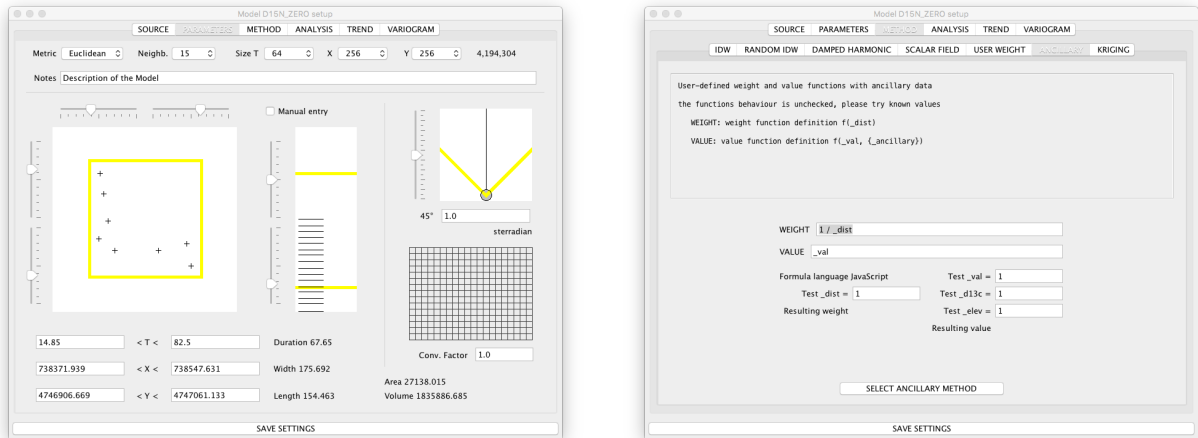
³⁹The dataset should be a *.csv* (comma separated values) text or file; each row representing a record, with its distinctive label (the id), space coordinates, time, value and, optionally, a few ancillary fields values.

similar parameters. The KILL function operates a clean deletion of the model from the database; care should be taken since there is no “undo” to deletion. The fact that the data are stored on an ordinary database means that they can be accessed straightforwardly within other application. **TimescapeLocal** uses a Hibernate mid-layer for accessing the data, so it can be attached to all supported database flavours.

The example data and the figures that follow are relative to the case study presented below (see section 4). The source dataset is a collection of mycorrhizal $\delta^{15}\text{N}$ from a symbiosis study.⁴⁰

3.1.2 The SETUP Panel

This section is the most delicate one. Users define their models’ parameters through a tabbed GUI⁴¹. A PARAMS and a MODEL panel allow the definition of all the relevant values and of the interpolation method.



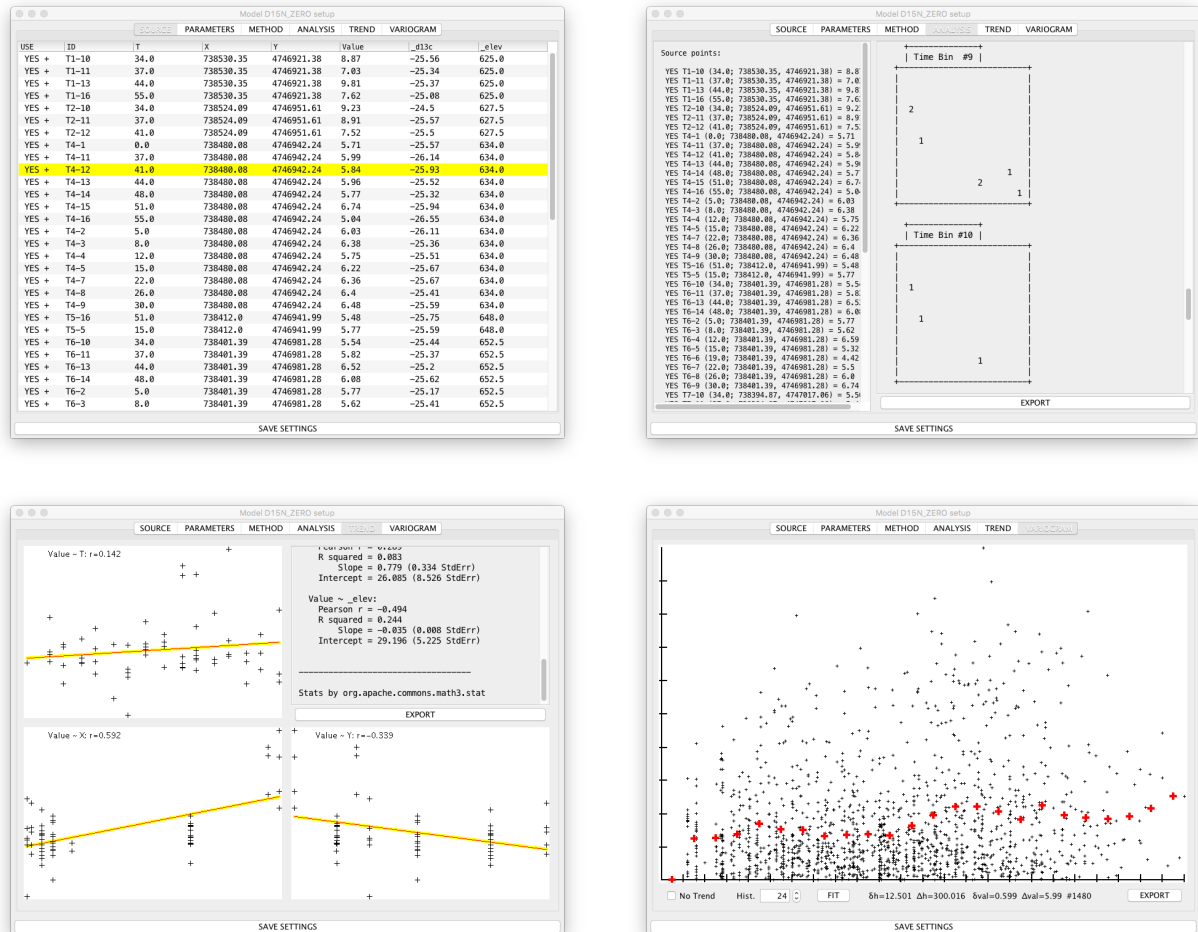
The parameters that should be defined can be categorised as follows:

- model consistency: the number of cells as width (x span) times height (y span) times the number of time sheets,
- boundaries: minimum and maximum bounds for space x and y , and time t coordinates,
- causality: the time to space conversion factor c and the causal cone aperture k ,
- neighbourhood: the number of near primes and the metric employed,
- method: the actual interpolation method.

⁴⁰The $\delta^{15}\text{N}$ is the relative difference, relative to a standard, of the ratio of the heavier ^{15}N isotope to the lighter and far more common ^{14}N . It is generally expressed in ‰ units for ease of reading. $\delta^{15}\text{N}$ is commonly employed in ecophysiology studies.

⁴¹Graphical User Interface

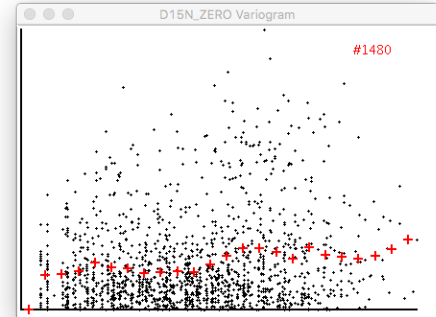
The **METHOD** panel is subdivided in a set of sub-panels, one for each method implemented. Users can also define their own methods, although this requires non trivial object programming skills. The methods include plain IDW (Inverse Distance Weighted) and some extensions, Kriging, and user-defined functions. Users can also switch off some sample points for advanced statistical testing. Users are assisted by a variety of statistical information about the source dataset:



- The **SOURCE** panel shows all the input dataset details. From this panel the single points can be switched on and off. Ancillary variables values, if present, are also shown here.
- The **ANALYSIS** panel shows a set of statistical analyses about the input dataset, including space and time distribution statistics and correlation analysis.
- the **TREND** panel shows a linear interpolation of the input values vs time and space coordinates (trend analysis). This is particularly useful for the trend removal in the variogram.
- The **VARIOGRAM**, as the name suggests, is the variogram plot. The variogram (and the other statistics) is updated whenever the user changes the value of the c add k parameters, or if a source point is switched.

The variogram is the most important piece of information the user is provided with. It is a spacetime variogram: it is computed with the couples of source points which are causally connected according to the value of the c and k parameters.

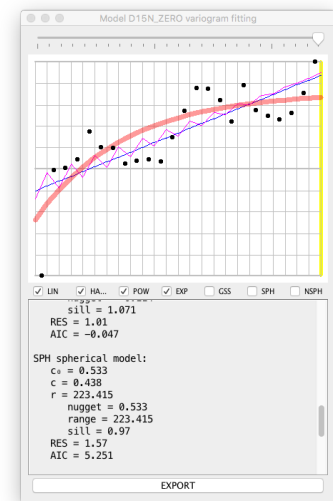
In the upper right corner of the window there is the number of point couples that affect the variogram; given the causal structure of the spacetime conversion, there is no need to divide the sums by two, as in ordinary variograms. User can control the number of bins displayed and the removal (or not) of the multilinear trend. All single points couples are shown (small black dots) as well as a bin-reduced (thick red crosses) version, showing the general behaviour. The general be-



haviour is so important that a small window is always open, so that users can see in real time the effect of changing the causal parameters values. The choice of the interpolation method has no effect on the variogram since it is computed only with the source points dataset.

The interpretation of the variogram is eased by the calculation of a few fitting functions: a linear and an harmonic fit,⁴² plus the traditional, Kriging-oriented Gaussian, exponential, spherical and double-spherical fits [Cressie 1990].

The values of the other statistics and trend analyses are not influenced by the c and k parameters, so their update is not so critical; they can change only if a source point is added or removed. The trend over time is especially important: the linear correlation can be negligible $R^2 \approx 0$ but a careful look at the values vs time regression plot can show a periodic component.



The IDW method can be adjusted for harmonic components choosing the DAMPED HARMONIC interpolator, the period should be tuned according to the pattern of change of the modelled phenomenon, most times it is one year, in forest ecology, where seasonality predominates. Users interested in multi-harmonic components should write themselves the code for the interpolator.

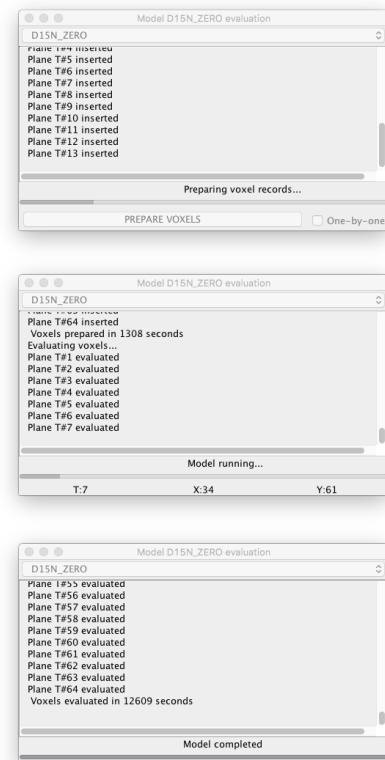
⁴²An harmonic fit is something unheard of in the realm of ordinary variograms, it reflects the presence of *seasonality*, which induces periodic (harmonic) oscillations of the sampled values. This is a common situation in many ecological studies and in forest ecology in particular, where seasonal variations play a key role.

3.1.3 Model Interpolation

The interpolation function has many customisation opportunities. Building up on the weighted sum idea, the evaluation of the weights is almost free. The most basic option is a simple weighted mean with the weights equal to the inverse of the distance, just a step above, we can consider the r th power of the distance: $w \sim 1/d^r$ maybe with a *mass* m :⁴³ $w \sim 1/(d^2 + m^2)^{\frac{r}{2}}$. All these functions are pre-packaged in the appropriate tabs of the **METHOD** panel. An **ANCILLARY** panel allows the user to define his/her interpolation weight functions, complementing the interpolation with ancillary values or even neglecting the measured values altogether, using the ancillaries alone. An harmonic weight function can be defined as well, as $w \sim \bar{w}(d) \sin(\omega \Delta t + \varphi_0)$, with a suitable frequency ω and a phase φ_0 , $\bar{w}$ is a decreasing function of the distance d .

The interpolation of the model goes on as follows, in a dedicated **RUN MODEL** panel:

- The elements of the model (voxels) are inserted into the database without evaluation (all the values are set to *null*). This step is important for the database safety, since it allows it to grow as fast as possible with a minimum of transaction activities (upper image).
- The voxels are evaluated one after the other, time sheet by time sheet. Each voxel value is then updated on the database. No table space growing happens in this phase (middle image). The evaluation can be parallelised easily, since there is no relationship among distinct voxels.⁴⁴
- Upon completion, a short report is shown, highlighting any relevant exception⁴⁵ that has occurred during the evaluation (lower image).



A progress bar in the lower portion of the panel shows graphically an estimate of the waiting time. Upon completion of each time sheet the user is informed about the time it took for its evaluation. It is important to check from time to time what's going on: an healthy database

⁴³The term *mass* derives from particle physics, where it corresponds to the mass of a scalar field. Here m^2 should be thought of as a dumping factor, a large m depresses the weight, while a small one has little effect on w ; the presence of m avoids the blow-up effect of w for the points located close to the source.

⁴⁴Unlike other kinds of three dimensional models, like the meteorological ones (e.g. weather forecast).

⁴⁵Technically, an *exception* does not necessarily mean an error condition.

would take the same time for the insertion of any sheet while, in the evaluation phase, the time intervals between sheets would grow, since the number of *null* voxels shrinks as time goes on.⁴⁶

The evaluation phase, for each voxel v , consists in a normalised summation of the values of the source point which fall inside the set of the possible causes of its central event:

- for each source event s , the distance d from v is $d = \sqrt{c^2(t_v - t_s)^2 + (x_v - x_s)^2 + (y_v - y_s)^2}$ or equivalently $d = \max \{c|t_v - t_s|, |x_v - x_s|, |y_v - y_s|\}$
- the associated weight is $w_s = \theta(kc|t_v - t_s| - \sqrt{(x_v - x_s)^2 + (y_v - y_s)^2}) / d$, the Heaviside θ function ensures that an s which is not causally connected with v has a zero weight.
- the associated value f_s is calculated. It can be simply the value of the source point or a more complex function
- the resulting voxel value is the weighted mean of all the values: $V = \left[\sum_s w_s \right]^{-1} \sum_s w_s f_s$.
If all the w_s are zero the voxel value remains *null*.

Many variations are possible on the scheme sketched above but all the relevant parameters are controlled in the **SETUP** phase, during the evaluation of the model there is no user interaction, nor it is possible to modify the parameters while a model is running.⁴⁷

The evaluation of a model is particularly core-stressing and the complexity allowed depends on the hardware capabilities. It is always safe to start with downscaled models before running a bulky model. The database activity is intense as well, in the insertion phase the storage space blows up fast but the transactional activity continues all over the run, with as many updates as the model's voxels.

As a rule of thumb, it is advisable to run at least a downscaled model not exceeding $100 \times 100 \times 100$ (one million) voxels before venturing the realm of billions, to see if the hardware is good for the job. Remind that doubling the space and time resolution of a model means an increase of almost $10\times$ of storage space and calculation time.

The cell spacing between voxels need not be alike, horizontal and vertical resolutions can be different and the number of time sheet has not to be necessarily comparable with the space resolution.

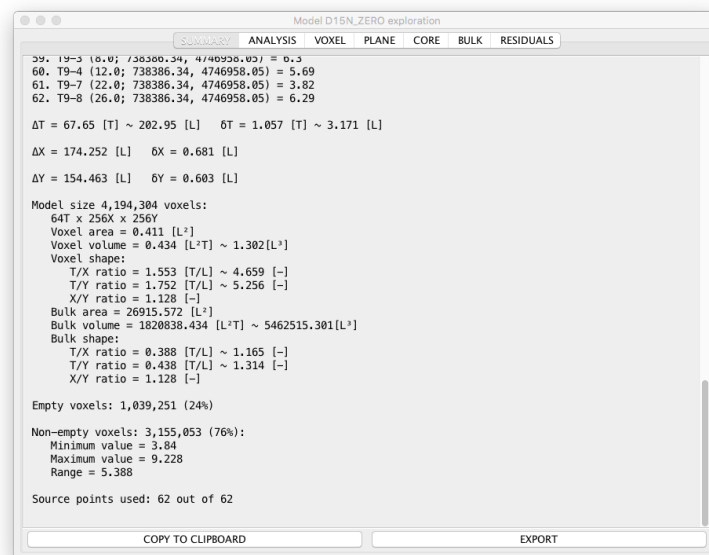
⁴⁶Null voxels are not updated, so no database interaction happens.

⁴⁷A running model is in a *locked* state that prevents any modification. If one needs a similar model with different parameters, it is possible to **CLONE** it, also if it is running.

3.1.4 The EXPLORATION Panel

A finished model is a collection of voxels in a single database table. A Timescape model does not have a “natural” way of displaying. The information contained in a Timescape has to be extracted somehow, according to the users’ needs. The **EXPLORATION** panel offers a set of statistical and graphical tools for examining the model’s values.

The statistical tools consist in three panels (**SUMMARY**, **ANALYSIS** and **RESIDUALS**), while the graphical tools are the most useful ones. It is also possible to export a variety of subsets from a Timescape, to be used as input layers in other GIS or statistical packages. For the sake of storage economy, the actual coordinates (space and time) are not stored in the data records. It is the *id* of the voxels (a triple of integer numbers) that can be used to recalculate the coordinates, so every export filter takes care of rebuilding the coordinates.

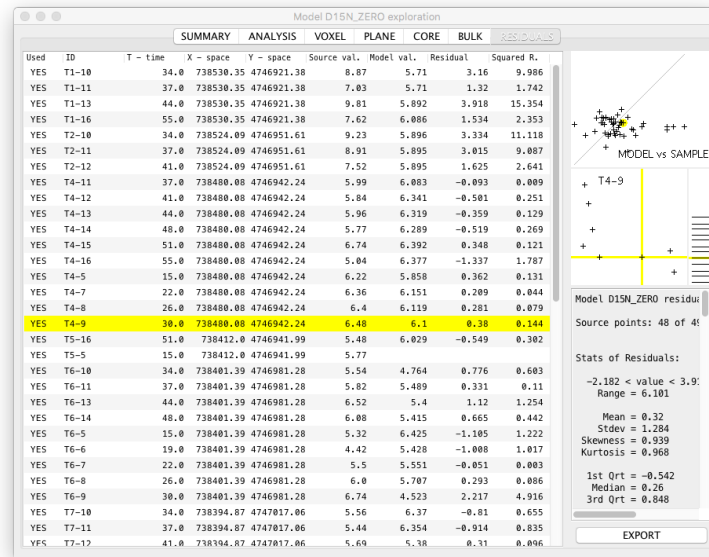


The **SUMMARY** panel

The **SUMMARY** panel shows a synthetic statistical analysis of the model. This report can be exported as a text file or copied to the clipboard for other uses. The informations shown include the geometry of the model (area and volume, both in LLT and LLL units), the number of *null* voxels and its ratio with respect to the total number. There are also some statistics about the values of the model.

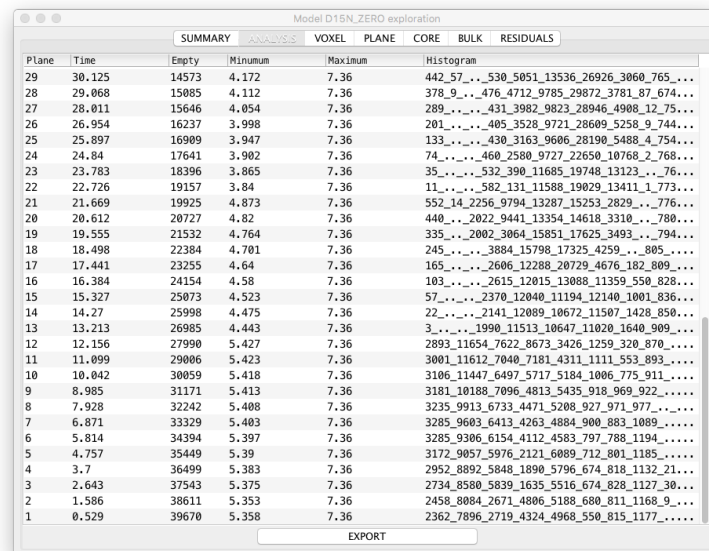
The **RESIDUALS** panel consists in the list of the source point, accompanied by the approximate value of the model in the same location (if it is included in the model extension) and the difference of these, or *residual*. It is also shown a QQ-plot of interpolated vs original values. Also this information can be exported, including the QQ-plot. A little map shows the position of the source points in space and time; clicking on a source point the map and the QQ-plot

points are highlighted, to let the user examine in detail the most interesting cases. Lastly, some statistics about the residuals are shown and exported as well.



The RESIDUALS panel

The ANALYSIS panel (below) consists in a sheet-by-sheet analysis of the values of the model. For each time the software calculates the minimum and maximum values, the number of *null* voxels and an histogram, represented as a string of occupation numbers.⁴⁸



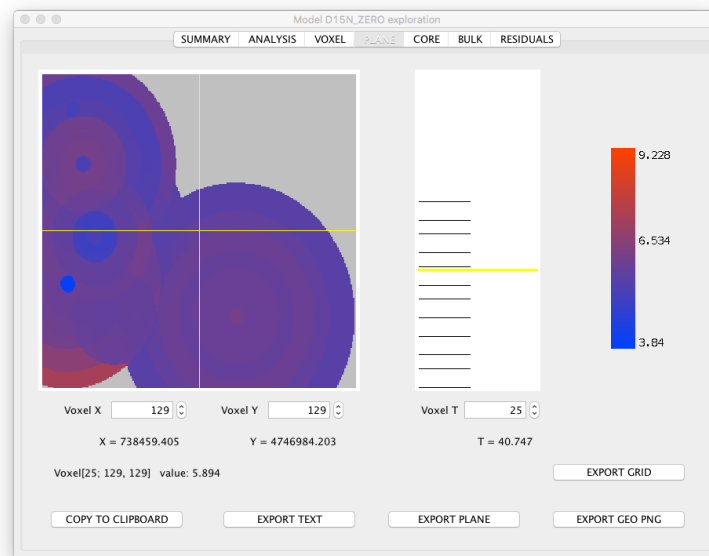
The ANALYSIS panel

Many ecological studies are focused on the variation over time of some quantity. This is the instrument of choice for finding the general behaviour and the extreme changes. Histograms are

⁴⁸The number of bins is configurable.

presented as strings of occupation numbers and not in graphical form due to the lack of space, but one can import such strings in any statistical software whenever a graphical representation is needed.

The most interesting panels are the four that bring a visual representation of some voxels subsets of the model. Basically, the model can be viewed as a cube, the base being the space extension and the height being the time interval. There are many ways in which such a cube can be sectioned, according to the kind of information that the user is looking for.



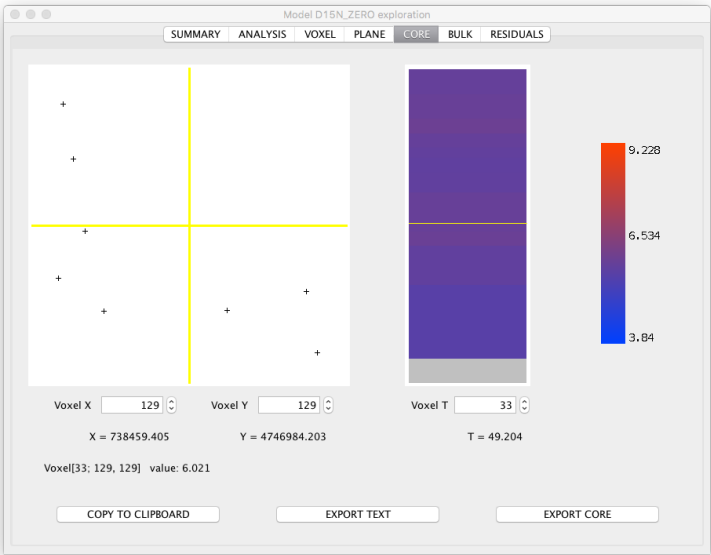
The PLANE panel

From a geostatistical point of view, the **PLANE** panel is the closest possible representation. Each plane of voxel, at a constant time, is an interpolated surface which can be shipped right to the user's GIS for further analyses. The surface, in fact, can be exported as an ascii grid of georeferenced values,⁴⁹ other export format include a georeferenced *.png* rendered image⁵⁰ and a formatted text file. The bar in the centre of the panel can be set to the time required; a set of horizontal lines show the time distribution of the samples dataset.

The **CORE** panel, on the other hand, is devoted to the site-wise analysis of the time evolution of the modelled values. The roles of the left and centre panels is exchanged: the left part shows a map of the area with the location of the samples and and hairline showing the position of the *core* dug into the voxels cube. This core is by any respect a time series of values, which can be exported as well as an ascii collection of values.

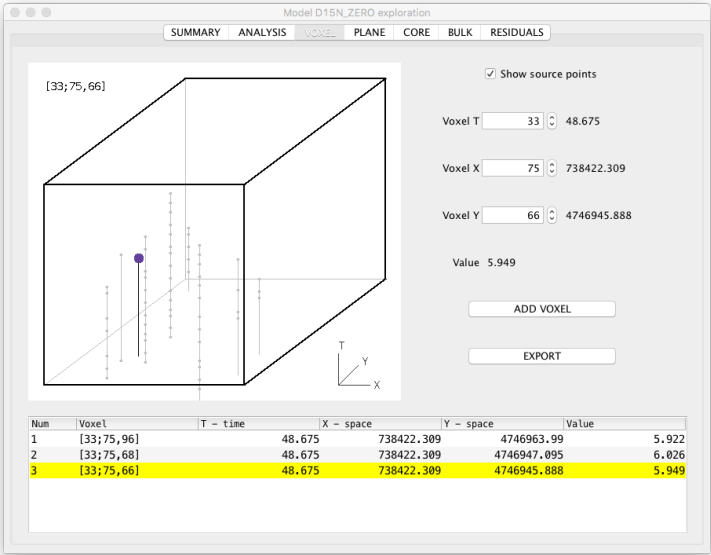
⁴⁹The format is the standard ESRI GRID.

⁵⁰A rendered image is only useful for representation in e.g. Google Earth, but it lack the actual values of the pixels.



The CORE panel

There is an endless literature about the treatment of time series (see e.g. [HYNDMAN]) in R and other statistical environments. This is where a Timescape can bridge the gap between ordinary “flat” geostatistics and time series analysis. Remember, however, that these time series are not measured values but they are interpolations, or even extrapolations if the maximum model time is greater than the maximum samples time.

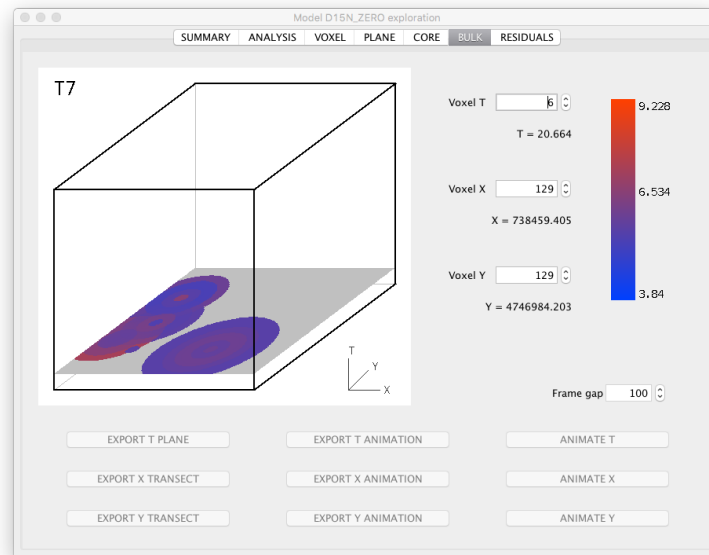


The VOXEL panel

The Timescape can be viewed as a cube and explored voxel-by-voxel in the VOXEL panel. This is really for a close inspection of small suspect areas. There is not a global vision of the values as planes or cores, so the user must know what he/she is looking for, and where. Users

can build their own table of voxels (the list in the bottom part of the panel) adding them one by one. This table can be exported in ascii format for further statistical investigations.

The main tool for having a sense of what's going on is the **BULK** tool. Also in this case the model is represented as a floating cube, that can be sliced in all directions.



The **BULK** panel

The model can be sliced horizontally, according constant time surfaces; this is basically what happens in the **PLANE** tool, too. We can also have constant- x and constant- y slices. These represent respectively the time evolution of North-South and East-West transects. A colour bar shows the values of the elements of the model (this legend is also shown in the other panels).

Having a global look at the model is important to find the hot spots and the hot momentum⁵¹ areas, this is what this tool has been designed. This panel is also the first place to see after the completion of a model, since it gives an intuitive graphic representation of the values.

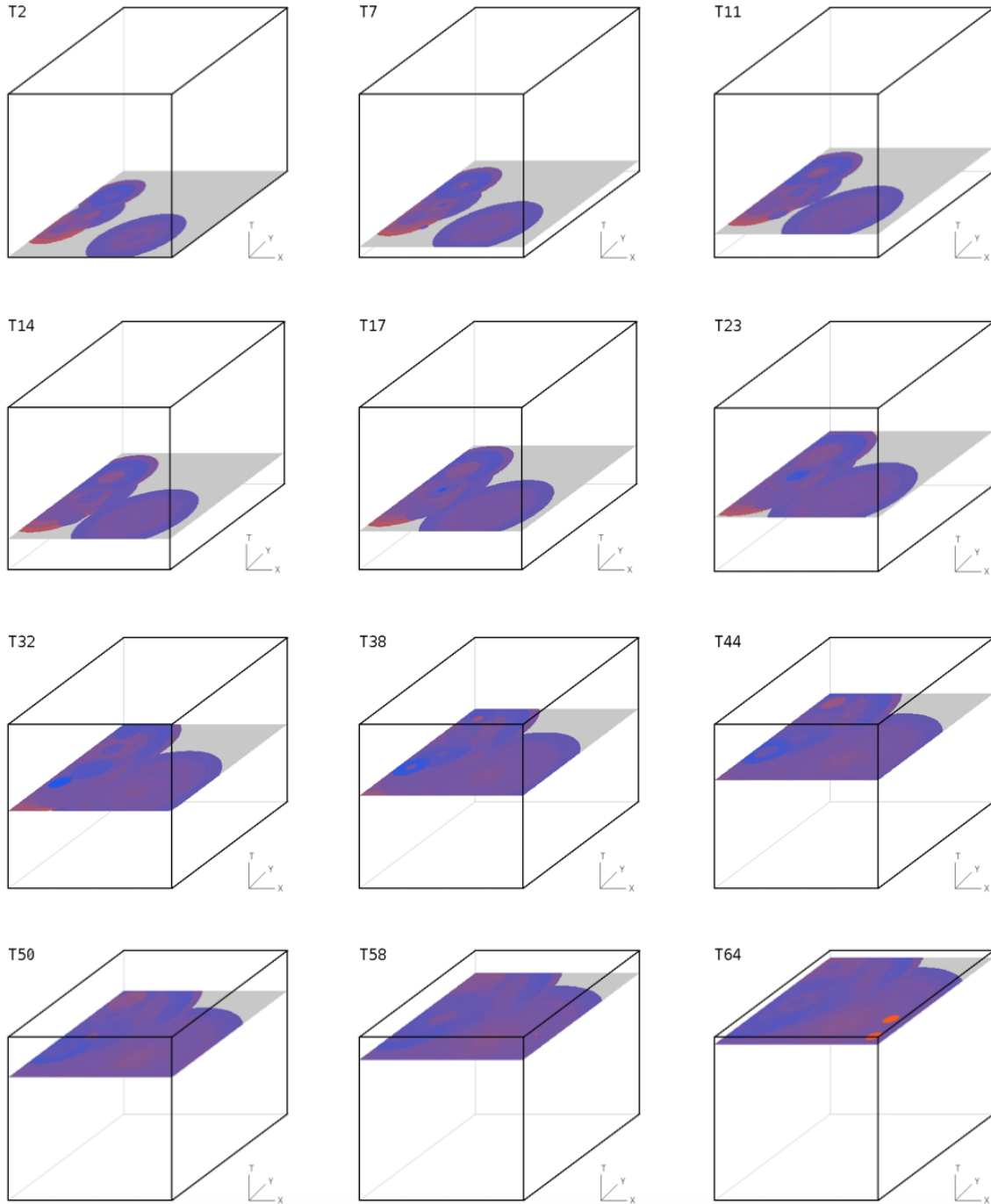
Users can select a value for x , y or t using the input sliders on the right side of the cube, so to have the appropriate plane shown, but it is the animation tool that is the most interesting one. Pressing **ANIMATE X**, **ANIMATE Y** or **ANIMATE T**, the planes are generated in sequence and plot one after the other. It is also possible to export these animations as *.gif* images. The animation procedure is very intensive both in terms of cores load and database accesses (all the elements of the model are selected in sequence), so bottom-of-the-line computers should be used with some care.

The following images show an example of animation along all the axes.⁵²

⁵¹Hot spots are relatively small areas of distinguishable values, non necessarily related to temperature, while hot momentum refers to small areas of rapidly changing values.

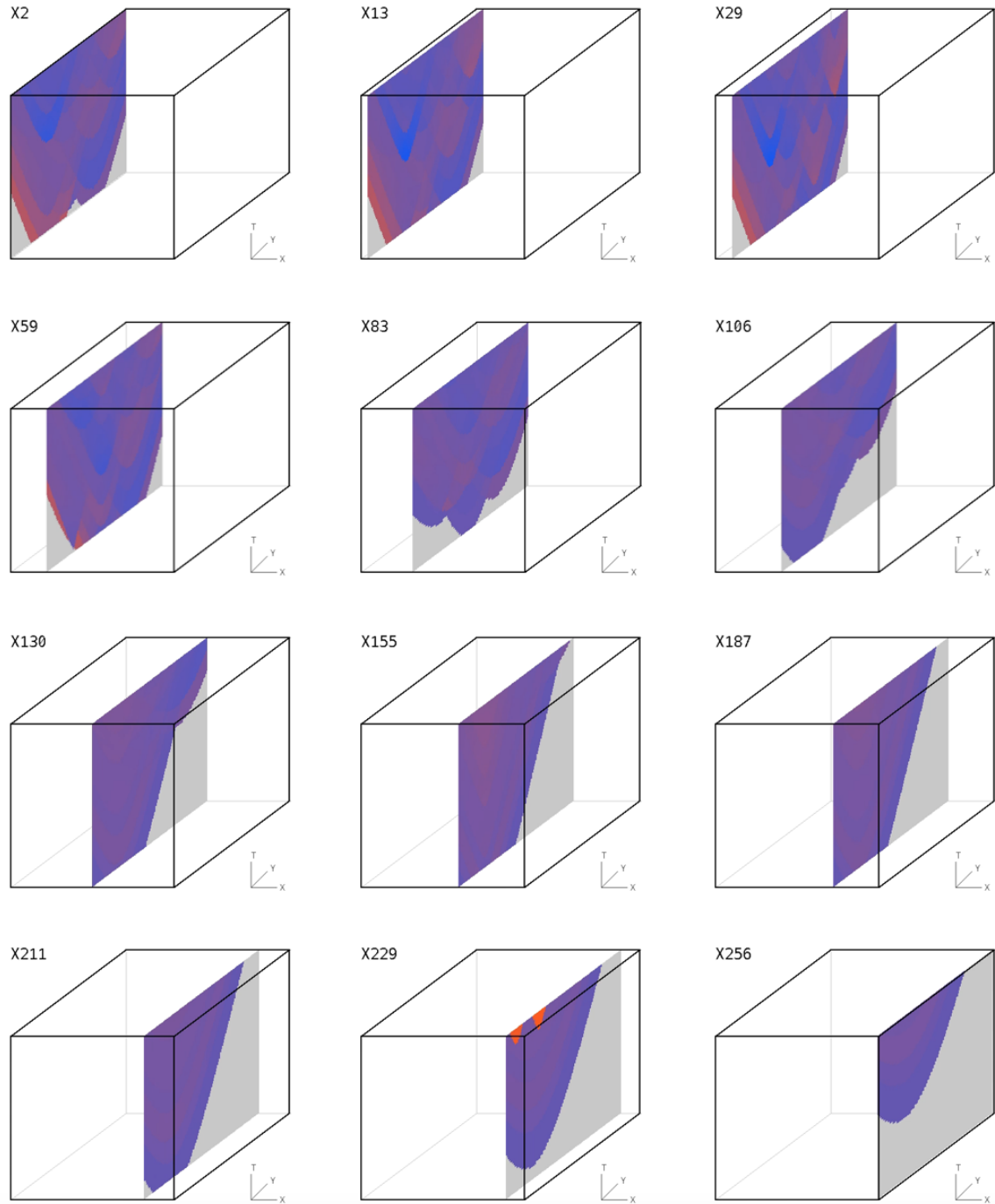
⁵²The Timescape is evaluated from a mycorrhizal $\delta^{15}\text{N}$ dataset (see section 4).

Constant T sections

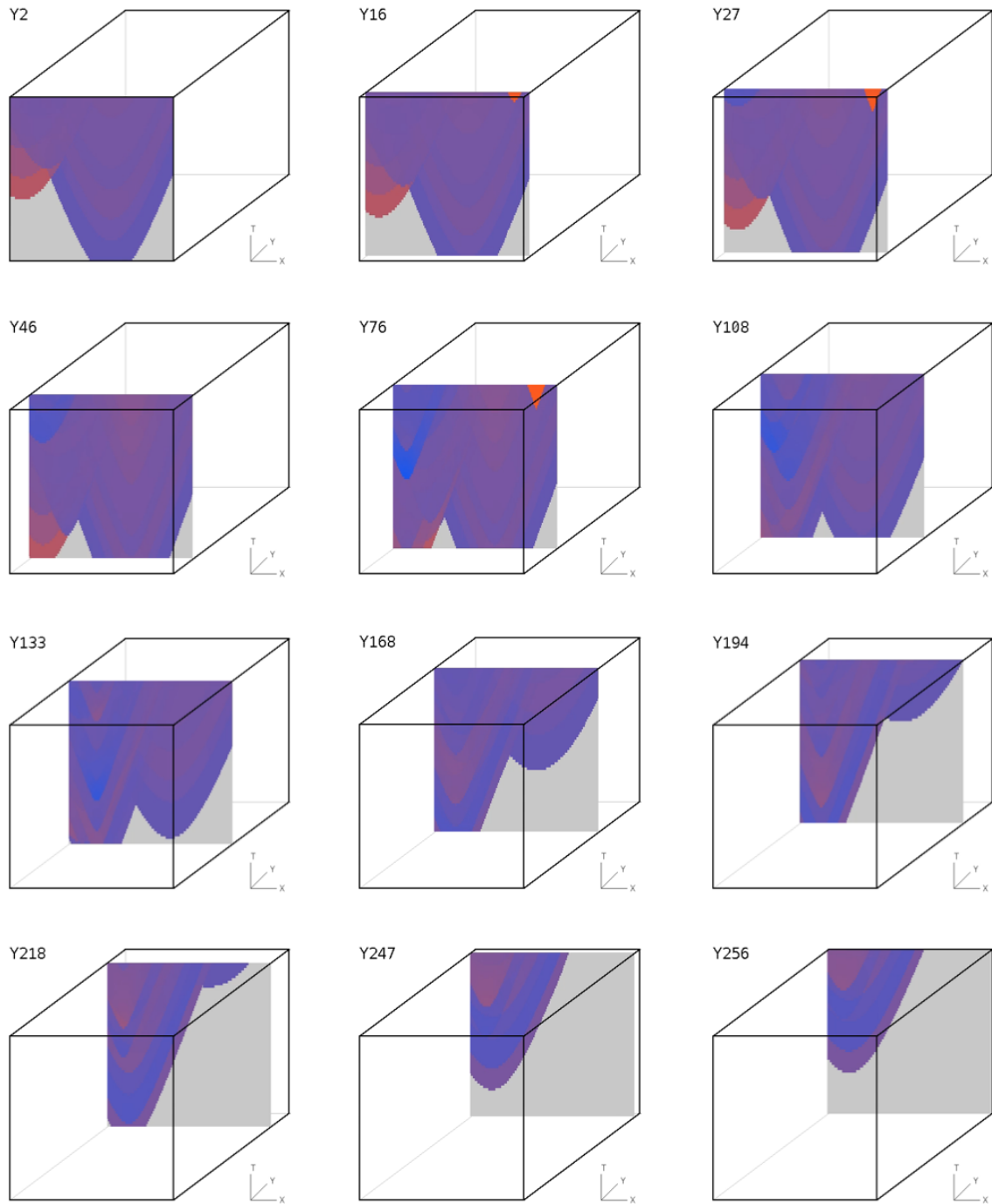


Time evolution of the spatial distribution of the values. Time increases bottom to top. See how the *null* areas shrink over time.

Constant X sections



Constant x transects from South (closer to the observer) to North, for different values of the horizontal coordinate. Any plane can be viewed as the time evolution of a transect. The effect of *knowledge increasing* in time is evident.

Constant Y sections

Constant y transects from East (left) to West, for different values of the vertical coordinate.

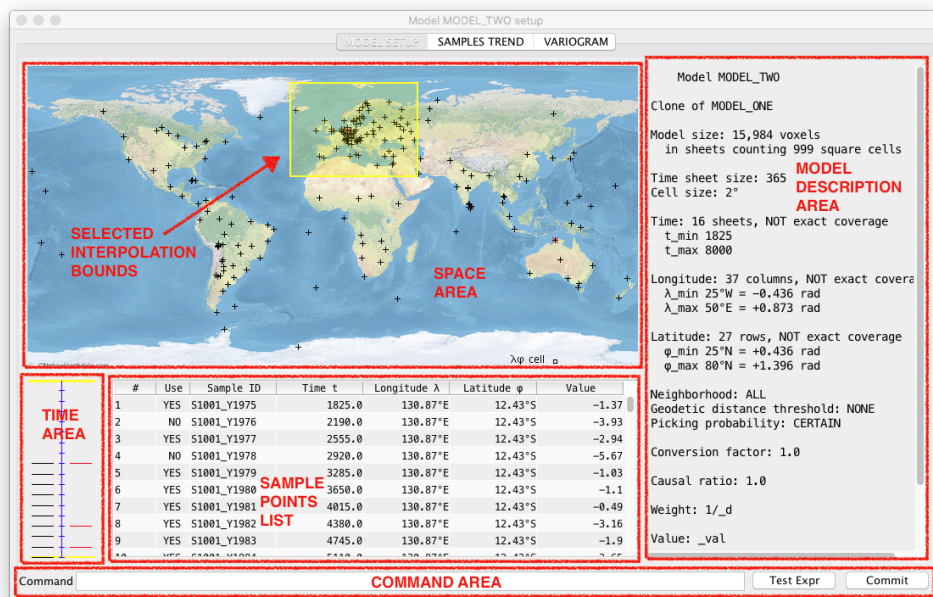
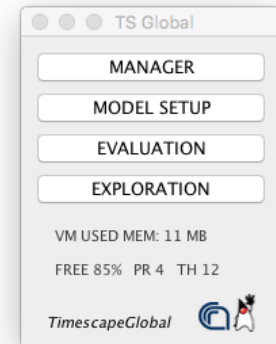
3.2 TimescapeGlobal

The global version of Timescape follows basically the same principles of the local one. The differences are related to the type of data involved, that in the Global case are spread worldwide, or at least on continent-sized area, which is untreatable in terms of projected coordinates.

The workflow of **TimescapeGlobal** follows the same steps of its Local counterpart: a model is created, its parameters defined, then it is evaluated. A Global finished model can be explored and exported in a few ways, with some differences and some limitations compared to the Local companion.

3.2.1 The SETUP Panel

The biggest difference is in the user interaction: the **SETUP** panel is designed about a command-line interface.⁵³ Users define their models' parameters through a few standard syntax commands (see the manual accompanying the software for any detail). The user interface window is subdivide according to functional areas. Unlike **TimescapeLocal** there are not active components but only text-based inputs. The *Command area* is the text area for writing such commands; the text is parsed pressing the *return* key and it turn red in case of a malformed syntax or an unacceptable parameter value.

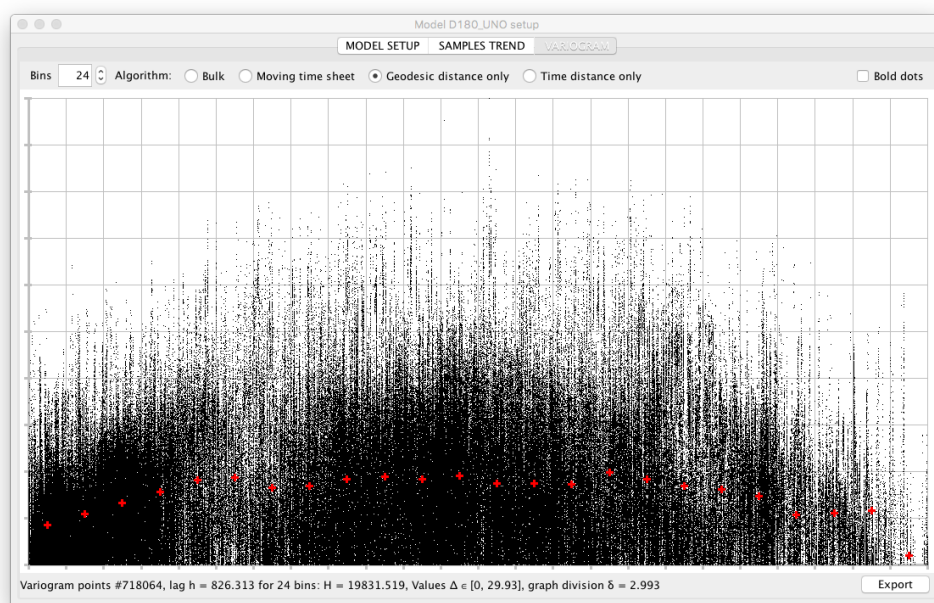


The SETUP panel

⁵³The **TimescapeLocal** development follows this user interaction paradigm too.

A world map⁵⁴ (the *Space area*) shows the distribution of the sample dataset, while their distribution over time is shown in the *Time area*. Between the map and the command area, there is a detailed list of all the samples, including their coordinates, time of collection and their value. Samples can be switched on and off for jackknifing and other sophisticated statistical techniques; by default a model is associated with the whole samples dataset.

The rightmost *Model Description area* shows all the relevant parameters of the model. Any change requested with a command⁵⁵ causes an immediate refresh of this area. The user is assisted statistically through two other panels, which show the variogram and a variety of statistics and trend analysis.



The VARIOGRAM panel

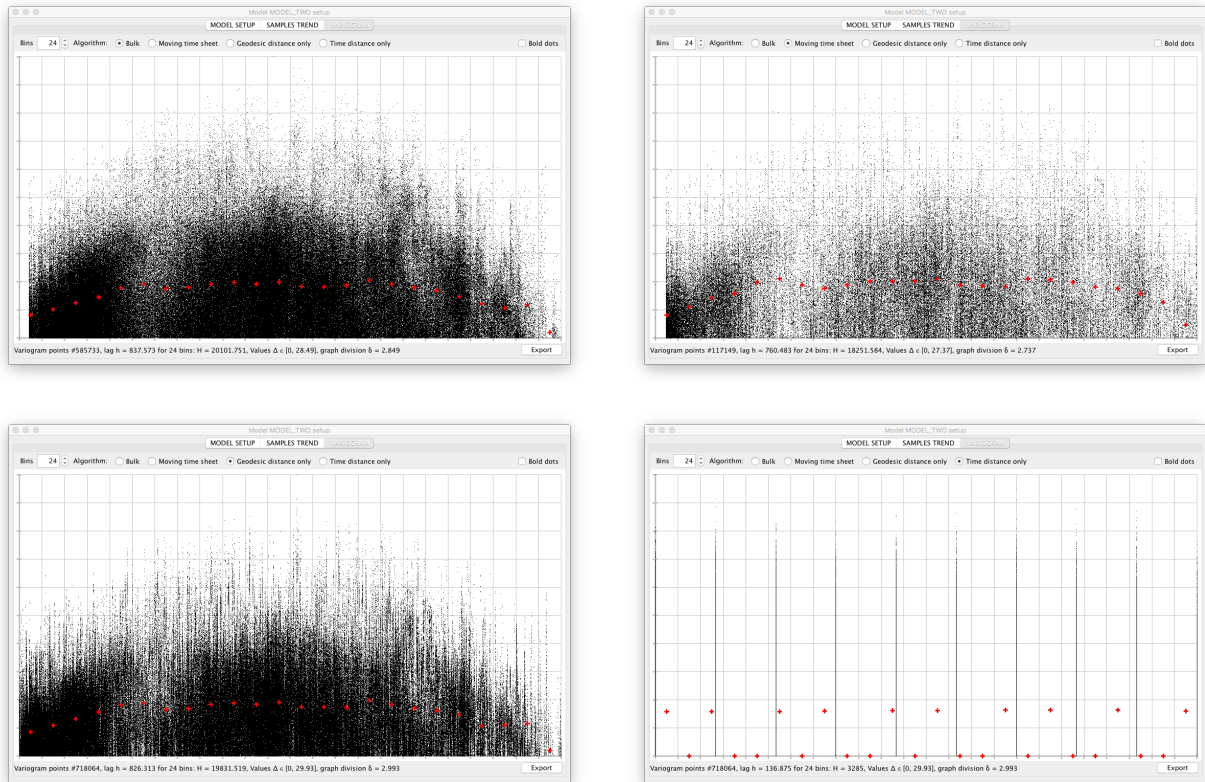
The variogram is especially complex and users have some customisation options to see the behaviour of the value vs different kinds of distances:

- **space** option: the distances are evaluated according to the geodesic arc length, regardless of the time of collection of the samples,
- **time** option: the opposite of the former, in this case only the time interval separating the couples of sample points is taken into account,
- **bulk** option: The distance of the variogram is exactly the one used in the evaluation of the model (this is the standard variogram), sometimes it is cluttered because of the number of constituent points,

⁵⁴The raster world map comes from [NATURALEARTHDATA].

⁵⁵The command syntax includes a Javascript-like function processor [ECMA].

- **moving sheet** option: this is a *moving window* version of the bulk variogram: for any point in the samples dataset, the variogram is calculated only with the other samples which lie within a model time resolution wide interval (it also means, automatically via the causal cone, that these points are also fairly close).

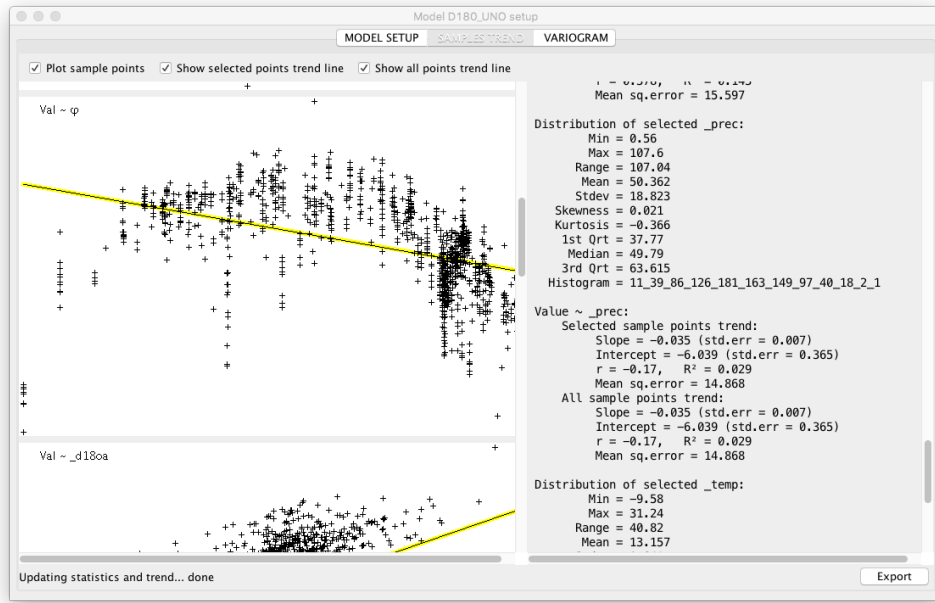


If the dataset sampling strategy was based on a fixed-time pattern, then the time-only variogram (bottom right panel in the above image) will be constituted of a series of spikes. The height of a spike is proportional to the variability among events, showing possible periodic oscillations. The moving sheet option (upper right) can be considered a decluttered version of the bulk variogram.

Using latitudinally extended datasets is often reflected in a somehow non-standard behaviour of the variogram. A standard, well-behaved variogram shows a nugget (an intrinsic variability also at a zero distance), an upward rising zone and finally an asymptotic plateau. Consider now some pairs of points more and more distant from the equator, but in different hemispheres, it can be the case that they have more or less similar values with increasing distances. In such a case, unlike the “standard” plateau, we can find a comparatively small decrease of the variogram trend for the largest distances. This is by no way a rare exotic phenomenon, think e.g. of the air or water temperature.

Another tool for helping in the parameters tuning procedure is offered in the **TREND** panel. This panel shows a series of linear and multilinear trend analyses: the values of the source

dataset are compared with the time of collection t , the longitude λ and the latitude φ .



The TREND panel

The values are also checked against any possible ancillary variable known. Multilinear analysis is performed, too: values as function of (λ, φ) and (t, λ, φ) .

The rightmost section of the TREND panel contains complementary text-based statistics, basically they are relative to the distribution of the samples' coordinates and ancillaries. All the standard statistics, the trend regressions (and other mathematical issues as far as possible) are evaluated by the Apache Commons Java libraries [APACHE].

TimescapeGlobal shows only linear regression analyses. A very low R^2 , however, means that there is no significant *linear* correlation, not that there is non correlation altogether. As the figure above shows, there is an evident correlation (the graph is relative to the rainfall stable isotope ratio $\delta^{18}O$ vs latitude φ). This is the trend analysis equivalent of the variogram decrease for large lags. These effects usually do not appear when the analysis is limited to a single hemisphere.

All the statistical analyses, trends and variogram plots can be saved for future reference.

3.2.2 The Model Evaluation

The model evaluation proceeds in two phases as in **TimescapeLocal**; the empty voxels are inserted before the actual calculation of their values. Also in the Global case every voxel is independent of all the others, so in principle any parallelisation of the evaluation can be done.

Compared to the Local version calculations, **TimescapeGlobal** is much more strenuous in terms of processor needs. On the other hand, the database requirements are about the same.

The evaluation goes as follows: for any voxel y are calculated the centre coordinates⁵⁶ $(t_y, \lambda_y, \varphi_y)$, then it is evaluated the contribution of each source point x , according to the weight

$$w(x) = \frac{\theta \left[k c (t_y - t_x) - R \arccos \left(\sin \varphi_x \sin \varphi_y + \cos \varphi_x \cos \varphi_y \cos(\lambda_x - \lambda_y) \right) \right]}{\sqrt{c^2 (t_x - t_y)^2 + R^2 \arccos^2 \left(\sin \varphi_x \sin \varphi_y + \cos \varphi_x \cos \varphi_y \cos(\lambda_x - \lambda_y) \right)}}$$

This is the simplest possible expression! The weight is just the inverse of the distance $\sqrt{\dots}$, but only if x is causally connected with y , that is controlled by the Heaviside step function⁵⁷ in the numerator. The distance is just the composition, though the Pythagorean theorem, of the space distance, evaluated along a spherical geodesic $R \arccos[\sin \varphi_x \sin \varphi_y + \cos \varphi_x \cos \varphi_y \cos(\lambda_x - \lambda_y)]$ with the “time distance” $c |t_x - t_y|$. Many variations are possible on this basic scheme, each of which adds further complexity to the calculation.

For any event x we then consider a contribution $v(x)$ which is, in the simplest case, just the value of the sample point, but it can be a more complicated function of the geometrical details and of the ancillary variables as well. For any detail refer to the maths in Appendix A.

The value v_y of the voxel y is the weighted average

$$v_y = \frac{\sum_x w(x) v(x)}{\sum_x w(x)}$$

if at least one of the the $w(x) \neq 0$, otherwise the value of y remains *null*.

During the evaluation it is created a report which contains, sheet after sheet, the running time and any error condition that has occurred while evaluating the sheet’s voxel values.⁵⁸

Compared to **TimescapeLocal**, the evaluation times are longer, but not too much. The very bottleneck of the software is the database so, unless the evaluation takes place on a really ill-equipped computer, expect no more than a doubling of the running times, for a similar number of voxels.

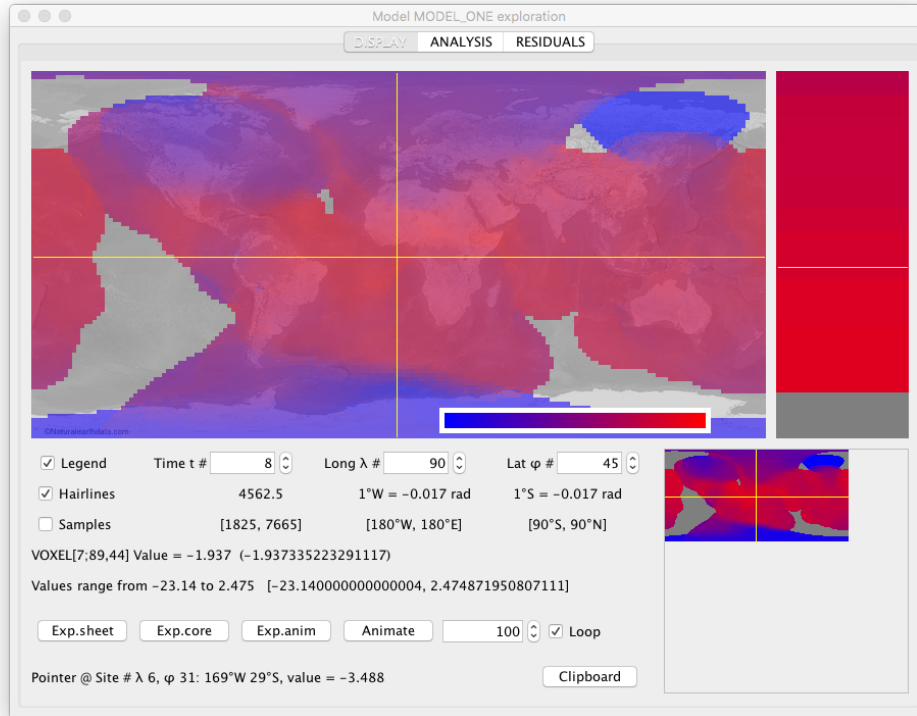
⁵⁶Since these coordinates are located on a regular spacetime lattice, they are calculated once and for all at the beginning of the evaluation.

⁵⁷The Heaviside θ is defined as $\theta(t) = 1$ if $t \geq 0$ and $\theta(\dots)$ and $\theta(t) = 0$ if $t < 0$. This function acts as an on/off switch, allowing only the x lying in the backwards causal cone of y to pass.

⁵⁸The simple formula sketched above is quite harmless and does not generate any error. If the user defines his/her own functions, these can generate runtime errors, which are logged by the software. The reading of the logs, however, requires some experience with Java programming.

3.2.3 The EXPLORATION Panel

A finished model can be explored in a few ways. Basically, we can extract GIS layers and time series. It is also possible to export an animation of the model as a whole as a *.gif* image.



The EXPLORATION panel

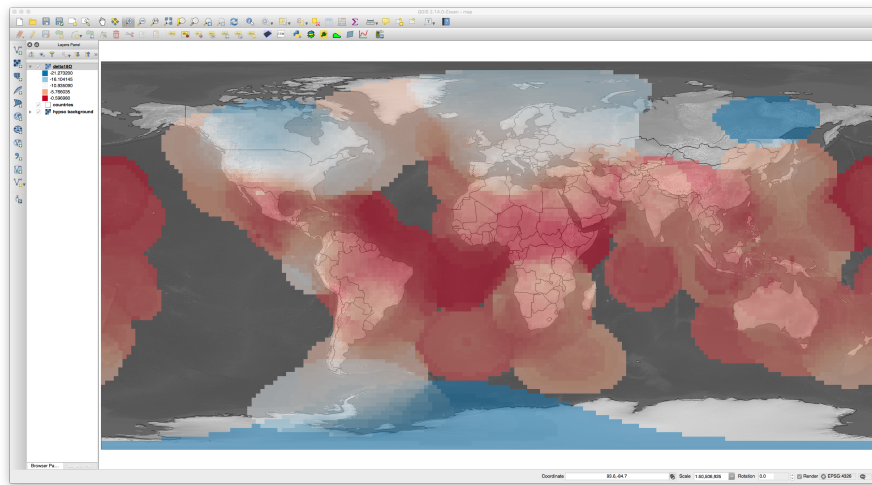
The EXPLORATION panel consists in three sections, the one above (DISPLAY) is the main arena for user interaction and for model exporting. The map (upper left) and core (upper right) sections are interlinked, so that clicking on a place shows up the corresponding core and clicking on a core sets the appropriate time on the map. A third image (lower right corner), also synchronised with the map, shows a one-to-one pixelwise model section. The map is superimposed with a slight transparency to a grayscale world image. The user can hide the colour scale and the yellow hairline, as well as the source points location in space and time.

Pressing the appropriate buttons users can export the corresponding data:

- An ESRI Grid at constant time, together with the world map in *.png* format,
- A time series at any given location (ascii file and *.png* image),
- An animation of the whole model⁵⁹ (see the image at the end of the chapter).

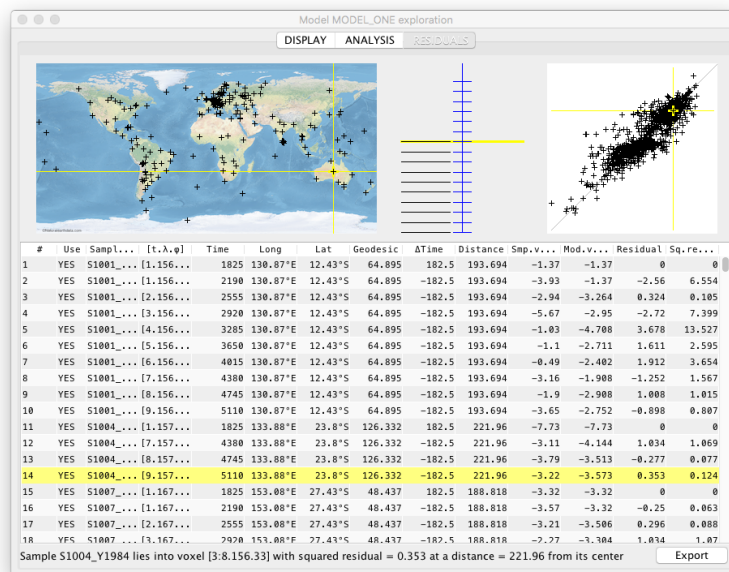
⁵⁹The animation, with exportation or not of the results, loads the processor and the database very heavily, since every equal-time sheet is shown in a fast sequence. This is exactly like **TimescapeLocal**, since the curvature of the Earth has no influence in the representation, which is in fact flat.

As the figure below shows, it is easy to import a Timescape equal-time grid in a GIS environment [QGIS], which is probably part of the standard workflow in a forest ecology study. Time series can be imported in a statistical environment like R [R, RSTUDIO].



A QGIS import

The exploration features include also an **ANALYSIS** and a **RESIDUAL** panel. Residuals are useful to check the model against the source values.

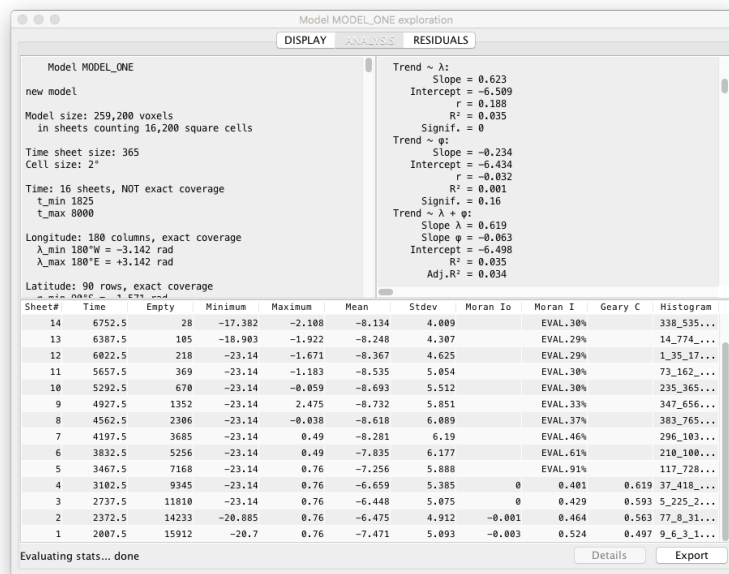


The **RESIDUAL** panel

The source dataset is ordered in a table on the lower portion of the panel. Clicking on a point's row highlights its position on the map, on the time panel and on the QQ-plot on the upper right corner. All these elements can be exported in ascii and image format.

The last panel shows some statistics: on the upper right corner there are the whole model statistics, trend analysis with respect to time and coordinates (linear models only) etc. The

main list shows sheet by sheet statistics, one row per constant time sheet. These include the analysis of distribution (minimum and maximum values, number of *null* voxels, histogram etc) and the autocorrelation estimators (Moran's I and Geary's C).⁶⁰

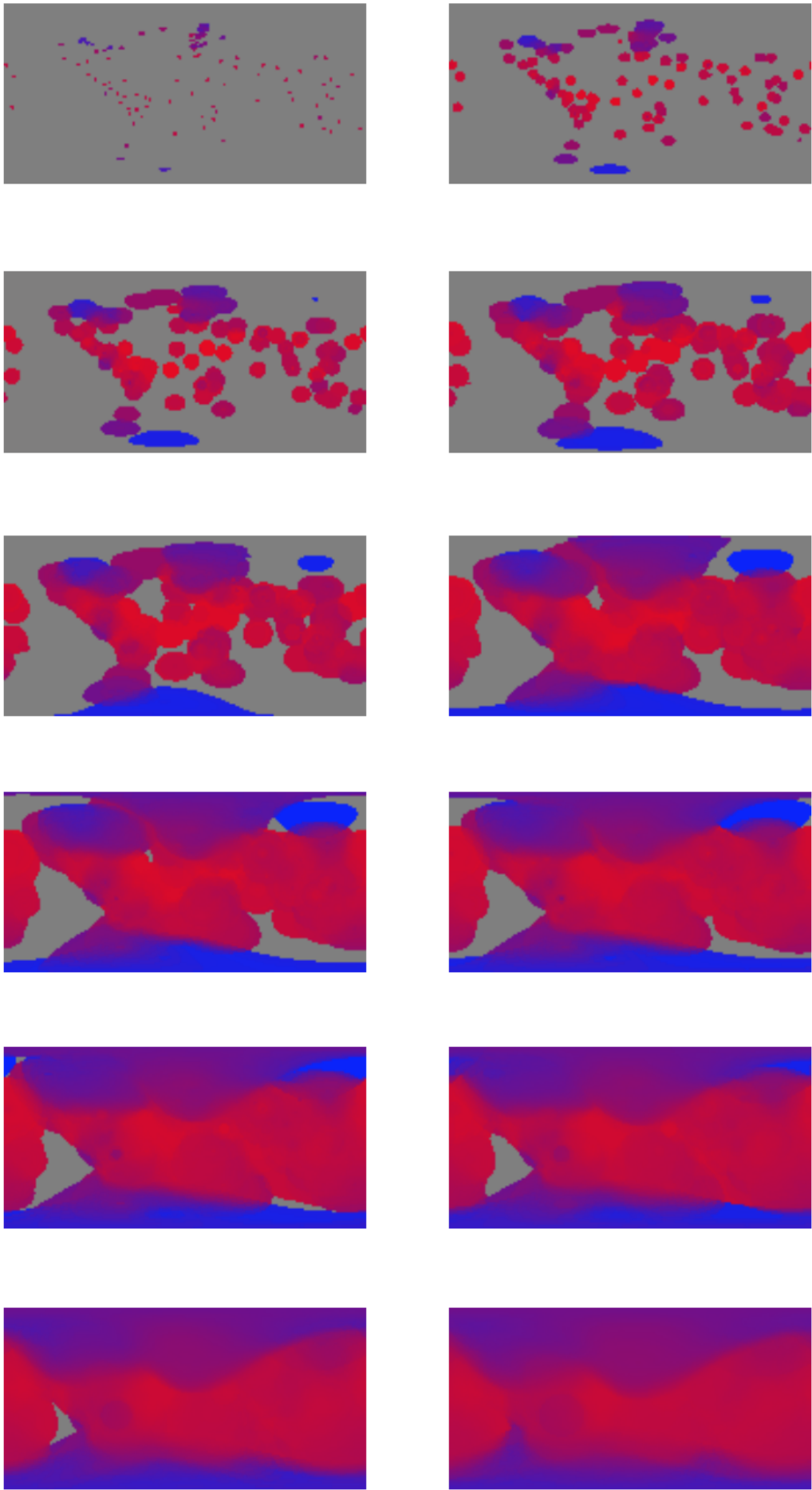


The ANALYSIS panel

At the opening of the panel the Moran and Geary columns are empty, instead of the indices there is a percentage of completion of the calculations. It is not surprising to find high autocorrelation values (not exactly a good clue in geostatistics) for very early times; this is due to a lack of information and the values tend to stabilise towards more acceptable figures for later times. The global linear models do not take into account the latitudinal effect so a low R^2 not necessarily means no correlation.

Also these statistical analyses can be exported, in ascii format, for further investigations.

⁶⁰The evaluation of Moran's I and Geary's C indices is painfully slow and processor-intensive. It is performed whenever a new model **EPLORATION** panel is opened.



A **TimescapeGlobal** animation. Time increases top to bottom, left to right.

4 Case Study: Mycorrhiza Survival Strategy

This study is centered on the relationships among mycorrhiza and host trees. It originated from a project of region Umbria about the productivity of black truffles in a thinned implanted pine woods, located in western Umbria, close to the boundaries with Lazio and Toscana regions. The area is located in a mid-mountain landscape, at an elevation of about 650 to 700 meters above sea level, within the Alleronia comune.

The landscape is characterised mainly by oak woods and pine woods, mostly implanted before 1960s, and a few cultivated areas. In the last fifteen years or so, some thinning operations have been carried on the pine woods, in order to restore the autochthonous vegetation.

All the area of study has been subjected to pine elimination at various degrees, from slight thinning to complete clearing in most of the area, during ten-twelve years. The sampling campaigns were conducted in the spring and summer of years 2012 and 2013, with most of data been collected in 2013, before and during the growing season of black truffles.

The black truffles (*Tuber aestivum*) production is of some importance in the local mountain economy as a sustainable complementary production; the thinning operations have somehow reduced the production of truffles. The scientific aim of the study was to find the mycorrhiza survival strategy to such disturbances to possibly help keep the truffles production to an useful amount. The fungi-tree relationship was investigated through stable isotopes methods.

The conclusion of the research is that there is still a symbiotic relationship among fungi and host pines, in the few spots where these are still alive, but that in the areas without standing pines the fungi are feeding on the remnant pine stumps, without moving to a new host. All the trees in the productive spots have been sampled. The research is focused on the spatial and temporal patterns of variability of the stable isotopes ratios (carbon ^{13}C and nitrogen ^{15}N).

In order to evaluate solid spatial statistics the sampling has been conducted by georeferencing all the samples on the filed during the collection.

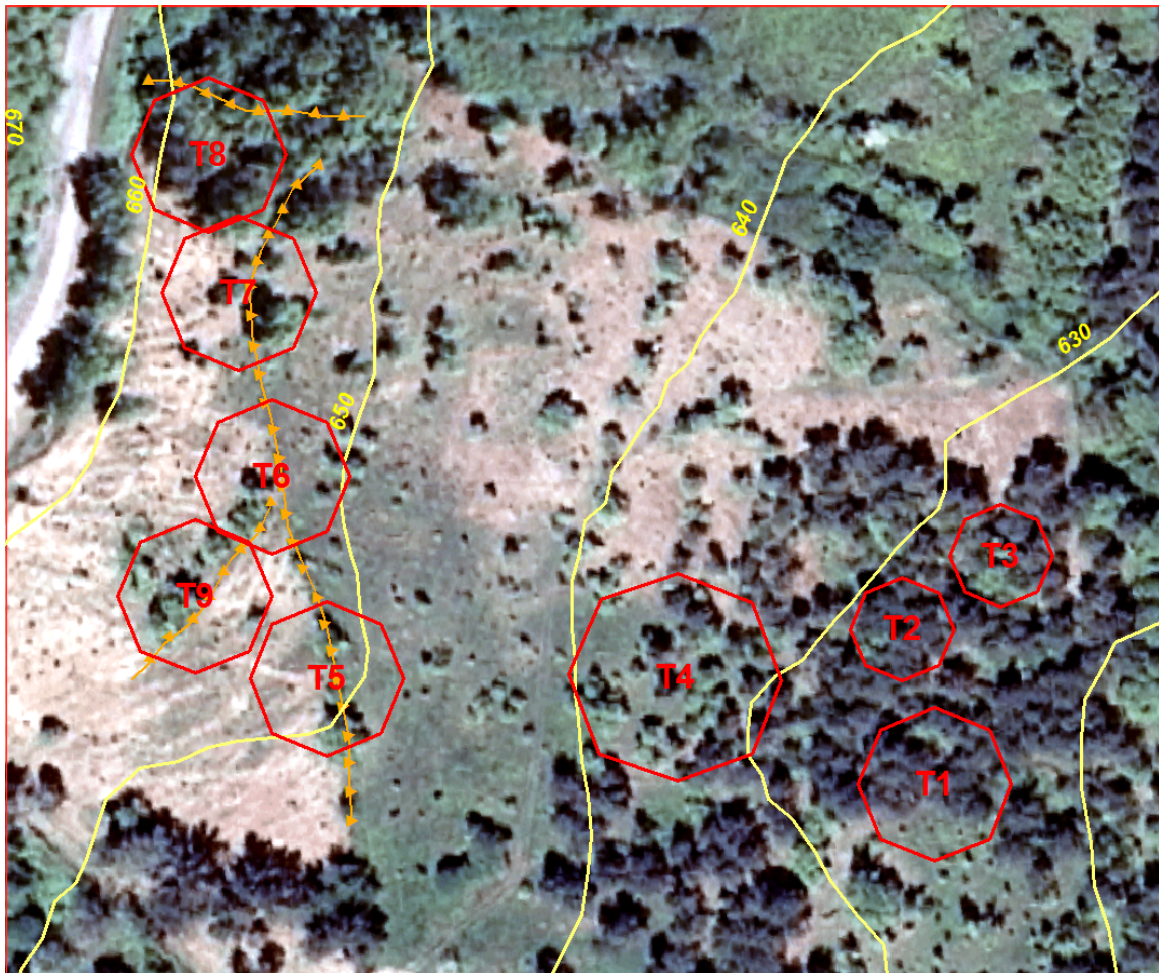
This study has been presented at the European Geospatial Union meeting [Ciolfi et al. 2016a] and at the First Italian Isotope Ratio Mass Spectrometry Day of the E. Mach Foundation [Ciolfi et al. 2016b].

4.1 Description of the Area

The area of study contains nine productive sites with different conditions in a relatively small space (few hundreds of meters across). The sites have been chosen for being the only known

productive ones within the area. Each site was enclosed in a circular area of variable radius.

The following map depicts the detailed topography of the area, which is roughly six hectares (300m E-W times 200m N-S). The contour lines (in yellow) vary from 620 to 670. Minor escarpments, about 1.5 meters tall, are shown in orange. There is an abandoned underground water pipeline crossing diagonally all the area (SW to NE) but it drains no water at all from many years so it has no effect on the hydrological behaviour of the area. Globally, the terrain is moderately steep with a difference of about 50 meters in elevation. The exposition is mostly E-SE. The elevated ground above the escarpment is probably the result of possibly ancient cultivations, also the positions of the bigger stones in this area seems to be of anthropic origin, though there is not evidence of dry stone walls.



The topography of the area over a 2011 orthophotograph

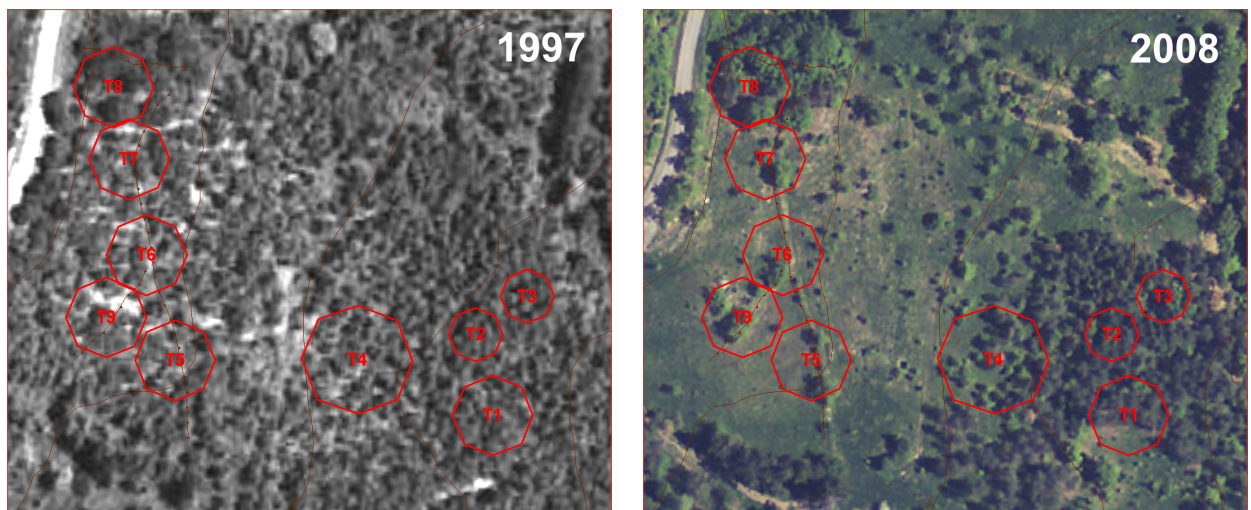
The nine sites of production vary considerably in coverage: some are almost untouched by the thinning operations, some are almost bare round and some thinned ones are been colonized by autochthonous trees.

In detail, the nine truffles-producing sites considered in the study are:

- T1: elevation 625m, flat terrain, standing pines covering most of the site, recent cuts.
- T2: elevation 628m, flat terrain, standing pines covering all the site, no truffles collected in the 2013 season, although it has been productive in 2012.
- T3: elevation 625m, flat terrain, all covered with pines.
- T4: elevation 630 to 640m, steep sloping terrain, mixed vegetation, oaks with few but big pines.
- T5: elevation 650m, includes an escarpment, sparse coverage, small trees, thinner soil compared to the other sites.
- T6: elevation 652m, rugged terrain including two escarpments, sparse coverage, small trees; possibly an ancient terrace.
- T7: elevation 655m, includes an escarpment, mainly covered by oaks and small trees.
- T8: elevation 660m, includes an escarpment and a dry ditch on the northern border, thick mixed vegetation (oaks and pines), close to the road.
- T9: elevation 652m, sparse vegetation, oaks and other small trees, includes an escarpment; possibly an ancient terrace.

The historical records of the cuttings are not completely known, anyway the pine coverage was complete in 1997, while in 2008 the situation was about like now. Thinning has gone over through recent cuts of single trees in the southeastern corner (sites T1 and T3).

The following images demonstrate the dramatic change in tree coverage.⁶¹



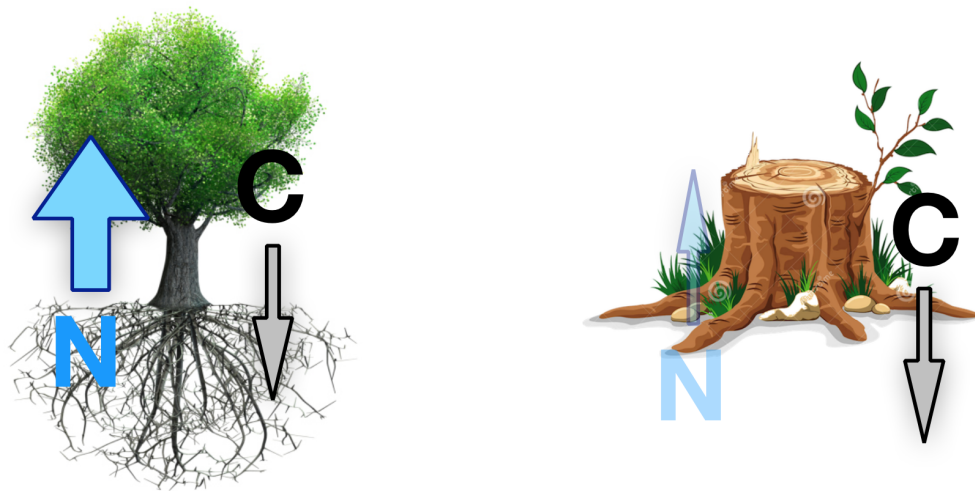
⁶¹The orthophotographs are taken from the WMS services of the Portale Cartografico Nazionale [PCN].

Appendix C contains more detailed information about the vegetation, including the position of any single tree.

4.2 The Symbiosis Model

The symbiotic relationships between mycorrhiza and host plant are complex and multifold [Deckmyn et al. 2014]: stable isotopes are known to be tracers of the photosynthetic activity of the tree leaves [Brugnoli and Farquhar 2000]. Carbon is taken by mycorrhiza from the host plant [Simard et al. 1997], the fractionation of carbon is low (not very selective). On the other hand, nitrogen compounds are transported from mycorrhiza to host tree, this process is quite selective so a large fractionation is expected in both ectomycorrhizal symbiosis [Höegberg et al. 1999, Werner and Schmidt 2002] and arbuscular mycorrhizal symbiosis [Govindarajulu et al. 2005]. Furthermore, a saprotrophic relationship is associated with no nitrogen fractionation [Hobbie et al. 2001] since it is a one-way relationship.

We employed a much simplified carbon-nitrogen exchange model which could be investigated through stable isotopes techniques.



The isotopic signature of a symbiotic relationship consists in low, if not negligible, differences in $\delta^{13}\text{C}$ and quite large differences in $\delta^{15}\text{N}$, up to 10‰, in terms of relative abundances.⁶² Small or negligible differences of $\delta^{15}\text{N}$ are expected among soil and truffles.⁶³ On the other hand, a saprophytic relationship, such as a fungi-stump one, misses entirely the nitrogen translocation from fungi to tree, so there is not a fractionation to look for [Henn and Chapela 2001].

⁶²See section 4.3.2 for an explanation of the delta notation.

⁶³In principle, there is a low fractionation from soil to fungi, but it is overcome by the fungi-tree subsequent fractionation

4.3 Materials and Methods

4.3.1 Sampling

The sampling has been conducted during the late spring and summer 2013. Trees leaves and soil have been collected in about one month, while truffles bodies were collected for about two months, from the beginning of July to the end of August.

The soil samples and the trees were georeferenced one by one, while the truffles were assigned the coordinates of the centres of the sites, for statistical consistency, since they were grouped by site. Some actual truffles collection points are reported on the detailed map in Appendix C. The position of truffles is affected by noise, however, since the finding is subjected to the collector's dog's nose and truffles stealing during the night is common, we have not used the quantitative estimate estimate of production, even if the masses of the truffles were measured. Collection occurred twice per week. A total of 330 truffles were collected.

For each site, the soil was collected radially in eight directions (N, NE, E, SE, S, SW, W and NW), along each arm we got four samples every five meters, plus the center. This sampling strategy is needed in order to spot any anisotropy of the values [Fortin and Dale 2011, Oliver and Webster 2015, Cressie 1990]. A total of 321 soil samples were collected.⁶⁴

Trees were individually georeferenced. The year's leaves were collected along three or four different directions, whenever possible, and mixed. Pines needles were collected following the same criteria, although the height of the trees often prevented the collection of more than two samples per tree. All the species were collected, including *Rosa canina* which is present in almost all the sites of the area. The most represented genus is *Quercus*, followed by *Pinus*. A total of 600 trees were sampled.

Georeferencing was critical for the spatial analyses to be conducted. The center coordinates of the nine sites were measured with two handheld GPS units (Garmin Montana 600t and Garmin Map 60csx equipped with Magellan Trailblazer XL external antenna). On each site center a spider's web graticule was build, along the four principal (North, East, South, West) and the intermediate (NE, SE, SW, NW) directions.⁶⁵ Plants and soil coordinates were sampled

⁶⁴A few more soli samples were collected out of the sites under investigation, since the sites distribution is skewed and the other samples could improve the reliability of the spatial models.

⁶⁵See the soil sampling map in the sampling atlas, Appendix C, page 99.

using polar coordinates relative to the sites' centres. A conservative estimate of precision sits below one metre.

The stumps and the actual truffles finding spots were collected at a later time and directly georeferenced through GPS measurement, with a slightly larger error, about two metres [Vischer 2006].

The resulting dataset has been stored in shapefile format [Goodchild et al 1992]. Raster images were acquired and stored in GeoTIFF format [GEOTIFF].

4.3.2 Isotopic Analyses

All the collected material was quickly dried within a few hours from collection at 70°C to stop any isotopic fractionation due to bacterial activity. The dried samples were finely ground. Leaves were analysed as one sample per tree, while three slices were taken from each truffle (before drying).

Stable isotope analyses of the samples were performed using an isotope ratio mass spectrometer (Isoprime, Cheadle, UK) equipped with a pyrolysis system (Euro Pyr-OH, Euro Vector Instruments & Software, Milan, Italy) and an elemental analyzer (NA1500, Carlo Erba, Milan, Italy), for measurements of $^{15}\text{N}/^{14}\text{N}$ and $^{13}\text{C}/^{12}\text{C}$ ratios.

The stable isotopes relative abundances were treated, as customary, as δ -values, given in ‰ units, referred to the corresponding standard:⁶⁶

$$\delta^n X = \frac{R - R_{std}}{R_{std}} \quad \text{i.e.} \quad \delta^n X = \frac{R}{R_{std}} - 1$$

where R is the heavier/lighter isotope ratio of the sample and R_{std} is referred to the standard.

Positive deltas are related to heavy-isotope enriched samples (with respect to the standard), while negative deltas are related to depleted samples [Slater et al. 2001].

4.3.3 Spatial Statistics

This study is centred on spatial statistical techniques. Standard statistical calculations were performed in the RStudio environment with the R statistical package [R, RSTUDIO], which is already equipped with many ecology packages [Borcard et al. 2011] and spatial features

⁶⁶Some authors use the notation $\delta^n X = \left(\frac{R - R_{std}}{R_{std}} \right) \times 1000$, which is somehow implied in the “given in ‰ units” sentence.

[Bivand et al. 2008]. GIS⁶⁷ raster interpolations were performed in the QGIS environment [QGIS], which also incorporates some GRASS functions [Neteler et al. 2011].

We followed basically the guidelines of Steininger for the usage of Open GIS Software in environmental sciences [Steiniger and Hay 2009, Steiniger and Hunter 2013], but for database flavour. We favoured MySQL [MySQL] against the more common PostgreSQL essentially for stability and portability, and because georeferencing and spatial indexing were not an issue [Urbano and Cagnacci 2014].

Most interpolation consist of standard geostatistical modeling, mainly ordinary Kriging [Cressie 1990]. In particular, Isoscapes techniques were employed [West et al. 2010]. An *Isoscape* is basically a spatial model of a stable isotope relative abundance, a sort of thematic map build over some sampled δ^nX . Isoscapes are a great source of information but require some care in their interpretation: it is somewhat implied that the driver of change is the position in space, it is an oversimplification which can be accepted most of times, if the isotopic relative abundances are stationary, i.e. if they are stable enough over time. This is the case, in this study, of the soil and the trees' leaves, but not of the truffles.

Some of the calculations involved the development of *ad hoc* techniques, which evolved to a full open source published software package, **Timescape**, which is also the object of this thesis.⁶⁸ Timescape techniques were employed for the analysis of truffles $\delta^{15}N$, which was unsatisfactory by standard geostatistical techniques.

The reason for mixing spatial statistics and time series analysis was that nitrogen fractionation occurred all over the collection time of the truffles (carbon fractionation is negligible), so it was impossible to compare samples from different days of collection.

The general Timescape algorithm has been presented at the European Geoscience Union General Assembly 2016 [Ciolfi et al. 2016a] and this stable isotopes application has been shown at the First Isotope Ratio Mass Spectrometry day [Ciolfi et al. 2016b].

Appendix A contains all the mathematical details of the algorithm. It is worth noting here that the Timescape interpolation technique is not weaker nor stronger than any ordinary (space-only) geostatistical procedure.⁶⁹

⁶⁷GIS stands for *Geographical Information System*, a collection of software (and hardware) resources for treating geographical issues.

⁶⁸Several versions of the software are available: **TimescapeZero**, the development version which was used for the truffles interpolations; **TimescapeGlobal**, which is available for download [TIMESCAPEGLOBAL] and **TimescapeLocal** which is available upon request to the author. **TimescapeZero** is the common ancestor of both the other versions, which differ for the kind of coordinates (global vs. local projected ones).

⁶⁹In a nutshell, the Timescape algorithms “reshuffles” and reorganises the coordinates, then it switches back to some standard interpolator method such as Kriging or IDW (Inverse Distance Weighting).

4.4 Results

The distribution of the samples was analysed spatially and isotopically.⁷⁰ The soil samples were picked according to a specific pattern for spatial interpolation (Kriging), so the resulting Isoscapes could be chosen as a reference for the remaining interpolations.

4.4.1 Ordinary Geostatistical Analysis

The leaves isotopic measurements (figure 1) show a distinctive carbon signature for pines at -27.0 ± 0.3 ‰ $\delta^{13}\text{C}$, while oaks are slightly lighter at -28.4 ± 0.1 ‰ with a very peaked distribution, which is somehow unexpected, since oaks varied a lot in terms of age, size and location. Incidentally *Rosa canina* shrubs show the same carbon signal as pines; their presence is ubiquitous all over the area.

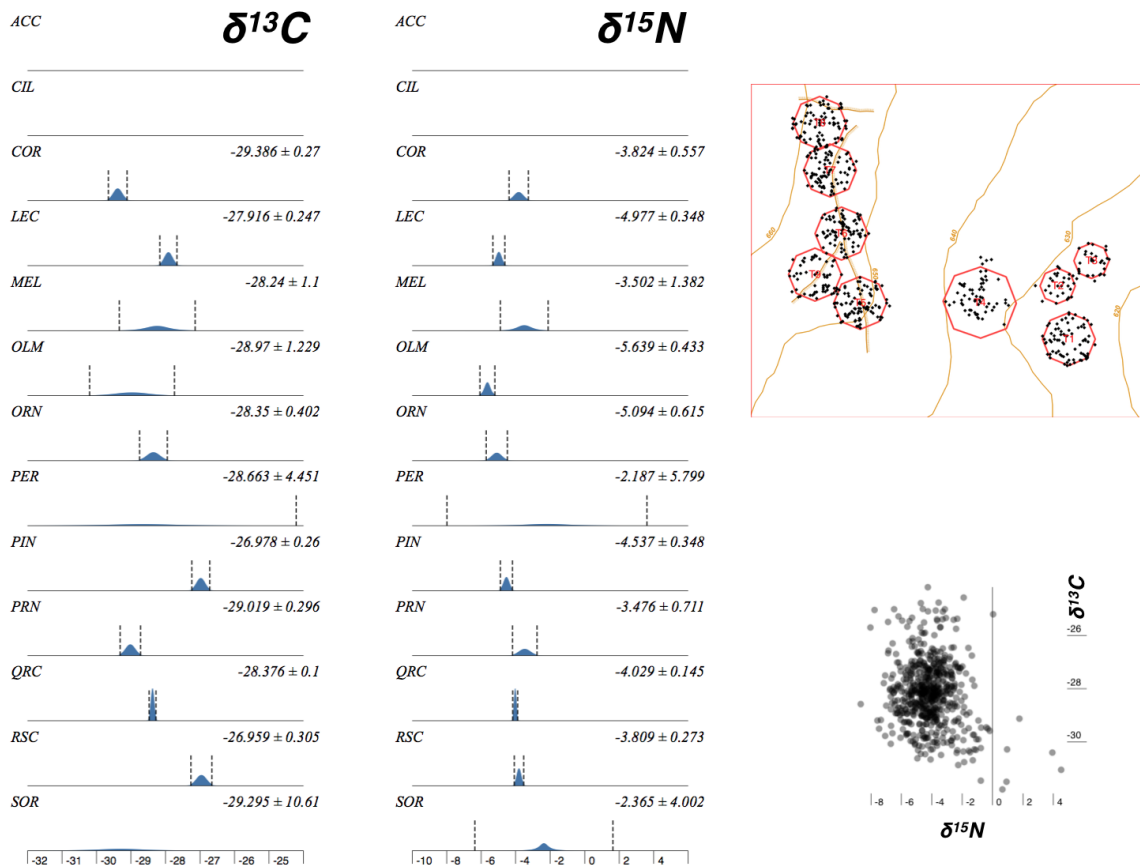


Figure 1: Leaves $\delta^{13}\text{C}$ (left) and $\delta^{15}\text{N}$ (middle) sorted by species: ACC *Acer campestre*, CIL *Prunus avium*, COR *Cornus mas*, LEC *Quercus ilex*, MEL *Malus sylvestris*, OLM *Ulmus spp*, ORN *Fraxinus ornus*, PER *Pyrus pyraister*, PIN *Pinus spp*, PRN *Prunus spp*, QRC *Quercus spp*, RSC *Rosa canina*, SOR *Sorbus aucuparia*. There is no significant cross-correlation among $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ (lower right). The spatial distribution is sketched in the upper right corner, without distinction of species.

Some species have too few representatives to make significant statistics and were not taken

⁷⁰More details on the samples positions can be found in the atlas in Appendix C.

into account for further investigations (*Malus sylvestris*, *Pyrus pyraister*, *Prunus avium* and *Sorbus aucuparia*).

Figure 2 shows the carbon and nitrogen relative abundances of the truffles, arranged by collection site. Carbon is not particularly interesting, while nitrogen presents significant differences among sites. In particular, only two of the sites show a value of about $+8.5\text{‰}$ $\delta^{15}\text{N}$ (T1 and T2), but the distributions are very broad, this is a clear sign of variability over time.

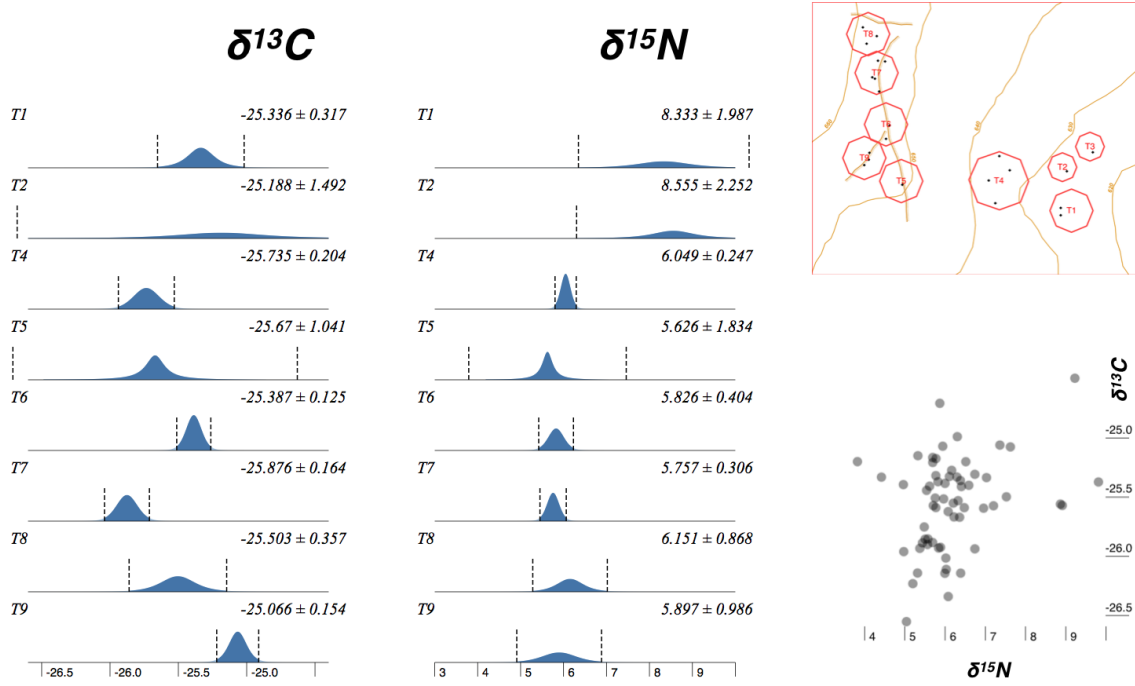


Figure 2: Fungi $\delta^{13}\text{C}$ (left) and $\delta^{15}\text{N}$ (middle) sorted by collection site; the values are averaged all over the collection time. There is no significant cross-correlation among $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ (lower right). The spatial distribution is sketched in the upper right corner.

Further investigations revealed that all the sites with pines had broad $\delta^{15}\text{N}$ distributions. This could be interpreted as a clue of symbiosis: ^{15}N fractionation is continuous in a symbiotic relationship: the fungi yield preferably ^{14}N , or better yet, the trees get preferably ^{14}N than ^{15}N , with the result that fungi get enriched in ^{15}N while the collection was going on [Hobbie and Colpaert 2003, Mayor et al 2009]. A single isoscape could non represent faithfully the distribution of $\delta^{15}\text{N}$ since the time variability had to be taken into account.

4.4.2 Timescape Analysis

A soil bias correction has been applied to the truffles isotope ratios because soil is the fungi nitrogen primary source substrate and some sites are significantly enriched in organic nitrogen (higher $\delta^{15}\text{N}$) than the others (figure 3). This correction, however, is a simple shift of the values, so it does not affect the shape of the distribution, which is exactly the same.

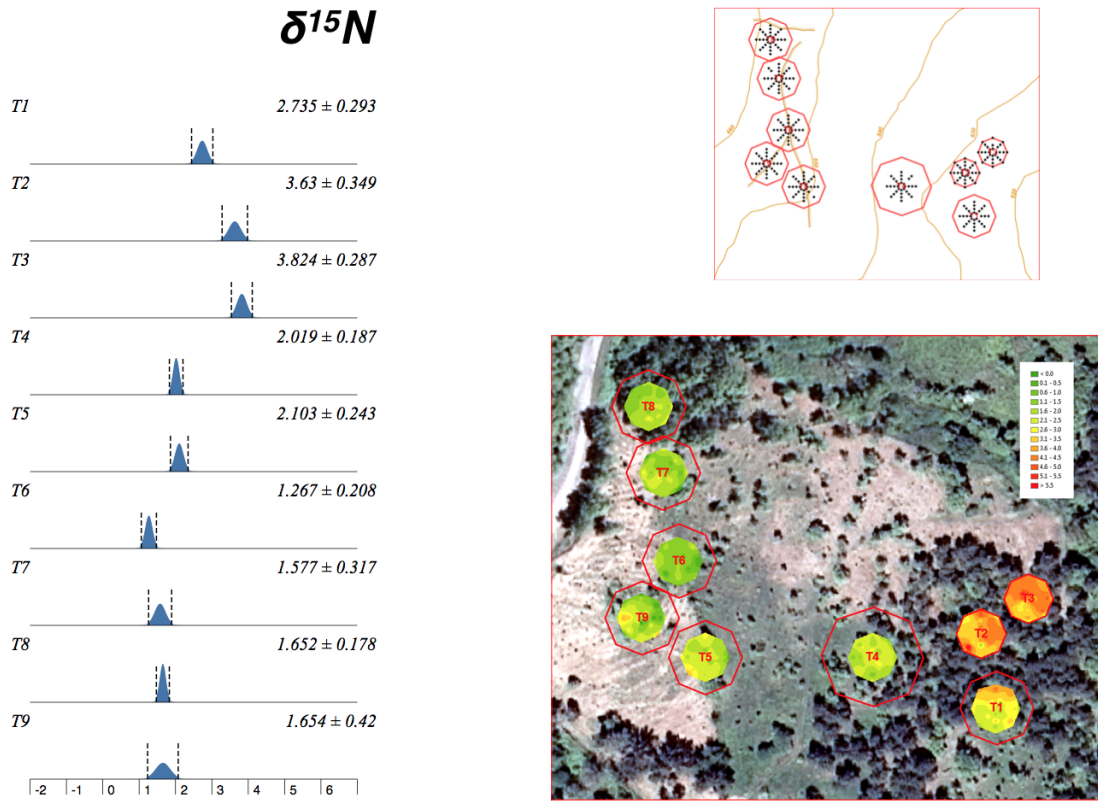


Figure 3: Soil $\delta^{15}\text{N}$ (left) sorted by collection site. The spatial distribution is sketched in the upper right corner and a full $\delta^{15}\text{N}$ isoscape is shown in the lower right.

The next and most important correction has been done through the extraction of time series of the values of $\delta^{15}\text{N}$ estimated by a Timescape model of fungi $\delta^{15}\text{N}$. As figure 4 shows we cannot take a single value of the truffles $\delta^{15}\text{N}$ per site. The figure shows some time series of truffles $\delta^{15}\text{N}$ from different sites. On the left hand side the time series show not significant variations from the average value, ranging $+5\text{‰}$ to $+6\text{‰}$ (see the corresponding peaked distributions on sites T4 to T7 in figure 2 above). On the right hand side the sites T1 and T2 show a broadened $\delta^{15}\text{N}$ distribution (figure 2).⁷¹ A careful look at the time series (figure 4) reveals a continuous ^{15}N enrichment over the time of collection (roughly two months): in this case the average is not representative of the actual nitrogen content of the fungi. The Timescape procedure allowed the estimate of a time-aware, corrected isoscape of fungi $\delta^{15}\text{N}$, to be cross-checked with the trees isotopic signatures.

Figure 4, other than the time-corrected isoscape, also shows the time evolution of a West-East transect (Y-Y) and a North-South transect (X-X). The red spots correspond to hot spot and hot momentum areas,⁷² a clear signature of enriched fungi and fractionation activity.

⁷¹Site T2 was not productive in the 2013 season.

⁷²Following the jargon of spatial statistics, *hot spots* are relatively small regions of clearly differentiated values, while *hot momentum* spots are small areas of rapid change of the values. Ordinary isoscapes can capture hotspots but not hot momentum regions.

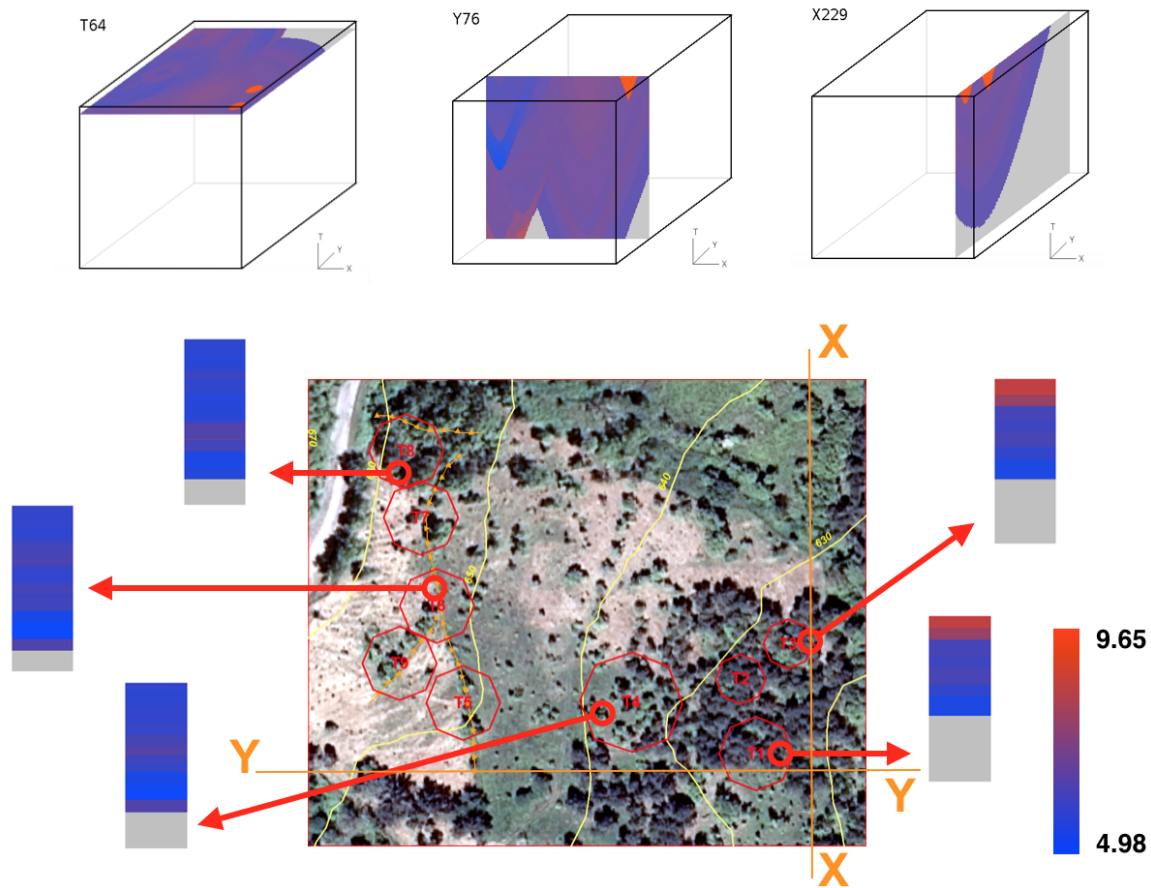


Figure 4: Fungi estimated $\delta^{15}\text{N}$ time series. on the left-hand sites (no standing pines) there are not significant changes of $\delta^{15}\text{N}$, while on the right-hand sites there is a clear ^{15}N enrichment over time, of about 3‰. The Timescape cube above is shown sectioned according to the corrected Isoscape (left) and according to the X-X (right) and Y-Y transects (middle). The red spots correspond to the latest, enriched truffles.

The statistical matching (high $\delta^{15}\text{N}$ difference and low $\delta^{13}\text{C}$ difference) of the soil bias and time corrected fungi isoscapes with the trees leaves isoscapes picked up only the pines as tentative hosts for mycorrhizal symbiosis. This could be enough for the T1, T2, T3 (lower right corner) and T8 (upper left corner) production sites, which are those with standing pines.

No significant statistical matching was found for the remaining sites. The truffles production, however, is not significantly different from site to site, suggesting a possible mycorrhiza survival strategy: switching to a saprophytic behaviour. It is possible that the fungi, without a living host, keep feeding on the remaining stumps. At the beginning of this study, for some reasons, stumps were not taken into account, due to the age of the thinning of the pine woods. Further sampling confirmed this clue, however. Pine stumps were found in all the pine-less productive sites (T4, T5, T6, T7 and T9) and a significant matching of the isotopic signatures was found.⁷³ The only obvious difference is in the nitrogen component, since there is no mycorrhiza to tree nitrogen transport, and consequently no fungi ^{15}N enrichment.

⁷³The actual location of the pine stumps dug in this study is portrayed in the map on page101, Appendix C.

The detailed distributions of stable isotopes ratios are reported in Appendix D. Every site has been treated independently from the others; the cumulative statistics of the site-averaged values⁷⁴ show a good agreement with the model, with high significance tests, in particular the ANOVA F-score and χ^2 (both $p < 0.001$) for soil $\delta^{15}\text{N}$ with respect to site, define a clear spatial differentiation. The truffles $\delta^{15}\text{N}$ also portrays really good tests, with an F-score's $p < 0.001$ and a slightly broader χ^2 at $p = 0.003$, still highly significant.

4.5 Discussion

Stable isotopes relative abundances are known tracers of physiological processes [Dawson et al. 2002]. Hydrogen and Oxygen are related to the water cycle and in general to the water use of the plants; in such a small plot, however we did not expect significant differences from site to site. Carbon and Nitrogen compounds exchanges from host plants to mycorrhiza and vice versa can be traced through the fractionation of ^{13}C (tree to mycorrhiza, low fractionation) and ^{15}N (mycorrhiza to tree, high fractionation). In fact, detailed studies of mycorrhizal relationship (e.g. [Lang et al 2013], which is focused on host tree genotype) show that complex spatial patterns emerge also in relatively small areas. The importance of mycorrhizal carbon input into soil organic matter is also reported in literature [Godbold et al 2006]. In this case we studied the symbiotic relationships through the geostatistical modeling of carbon and nitrogen stable isotopes from the sampled tissues.

Exploiting the correlations from the spatial distribution of carbon ($\delta^{13}\text{C}$ isoscape) and nitrogen ($\delta^{15}\text{N}$ time-corrected isoscape), according to the low carbon 13 fractionation plus high nitrogen 15 fractionation model, there emerge three distinct zones, outlined in Figure 5, which show synthetically the results of the study. It is possible to individuate clearly a symbiotic relationship only for the less disturbed areas, still populated by pines, corresponding to the T1, T2, T3 and T8 productive sites.

This spatial pattern matches the actual pines coverage of the area. The supposedly saprophytic area. The stumps dug from this area⁷⁵ match the model too, with a distinctive lower difference of $\delta^{15}\text{N} \sim 4\text{‰}$ against a 14‰ attributed to the symbiotic case [Peterson and Fry 1987]. The SAP area bears the signs of old cultivations (much before the pine implants of the mid-1900s) [Olivera et al 2011] and of a dismissed underground water pipe (now completely dry).

The details of the isotopic values are reported in figure 6, which is plotted in the abstract $\delta^{15}\text{N}$ vs $\delta^{13}\text{C}$ space. The truffles are well clustered at about -26 to -25‰ $\delta^{13}\text{C}$ and $+4$ to $+10$

⁷⁴Taking into account the Timescape correction for the mycorrhizal nitrogen.

⁷⁵Appendix C contains a detailed map of the stumps (page 101).

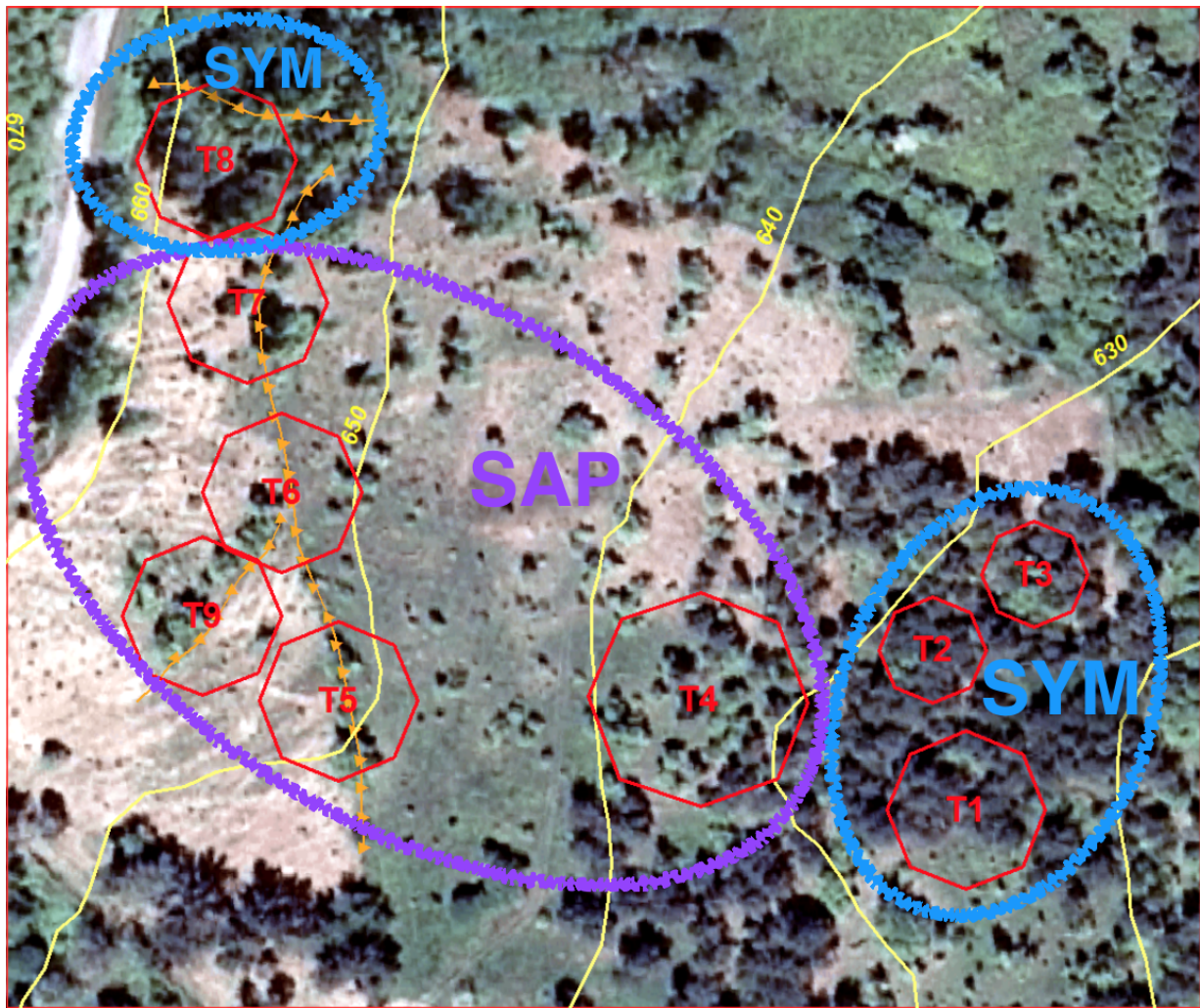


Figure 5: The cross-correlation of the isotopic signatures of *T. aestivum* and trees leaves outline a symbiotic area (SYM), where the living pines are still the dominating species, and a possibly saprophytic larger area (SAP). The upper-left corner SYM area, which is covered by pines, although characterised by different microclimatic conditions, shows the same symbiotic signature of the lower right areas.

‰ $\delta^{15}\text{N}$. The blue truffles subcluster⁷⁶ is referred to the symbiotic group, which host living pines are outlined as well; the map on the upper left corner highlights the interested sites.

The points between the blue and the purple clusters correspond to earlier measurements of the same symbiotic mycorrhiza [Hobbie and Colpaert 2003].

The purple subclusters are referred to the fungi that probably survive feeding on the pine stumps [Mayor et al 2009, Hobbie et al. 2001]. The trees have been cut at soil level, so that the stumps were almost unnoticed during the leaves sampling campaign and they were not taken into account from the very beginning of the study.

⁷⁶The subclusters are more populated than it seems from the figure, since there are many almost identical that give rise to values superimposed dots in the graph.

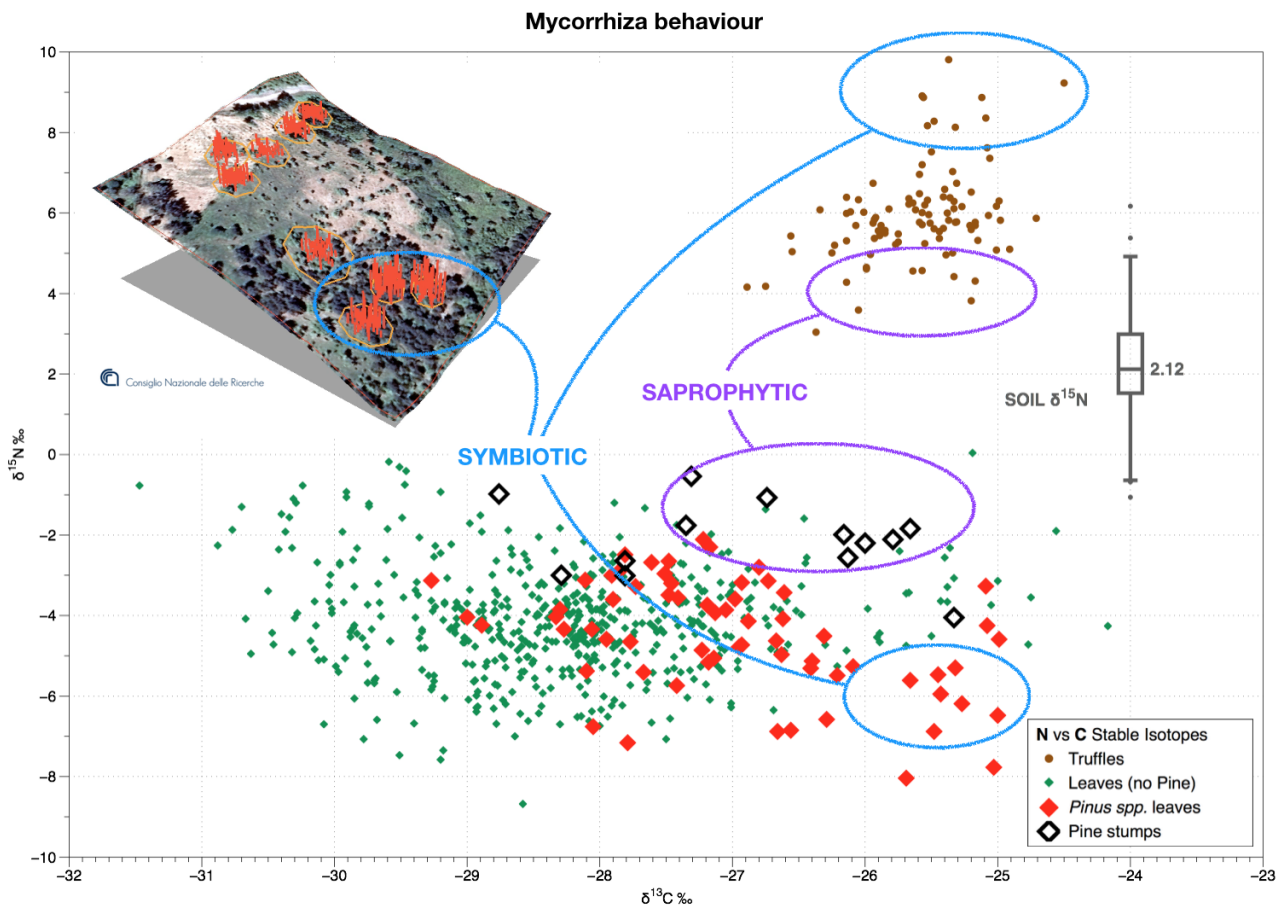


Figure 6: Modelled mycorrhiza behaviour according to the isotopic signatures. The height of the red spikes (one per tree) in the map is proportional to the probability of symbiosis. This is a synthetic representation, detailed site-by-site statistics are reported in Appendix D.

4.5.1 Conclusion: Complex Modeling for Complex Environments

This study demonstrates the power of spatial statistics in ecology studies. The time variation of the sampled data had to be taken into account, too, leading to the development of a spatiotemporal integrated approach. The resulting algorithm, **Timescape**, has already been published in the form of a freely-available, open source package.⁷⁷ The source code is part of the distribution.

Forests are very complex systems; from the modeling point of view, they are open to any sort of external influence, so modeling always requires some compromise. Many times the subtleties of statistics are overwhelmed by countless sources of disturbance that hides the spatial patterns of the variables of interest but sometimes, taking also the temporal variability into account, these patterns emerge clearly, as this case of the continuously fractionating *Tuber aestivum* demonstrates.

⁷⁷The download is hosted by SourceForge, see [TIMESCAPEGLOBAL].

5 Case Study: Olive Oil Provenance Assessment

The authentication of the geographical origin of food commodities is increasingly important in the food production and distribution industry. This study relates the spatial distribution of carbon and oxygen stable isotopes ($\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ isoscapes) of extra virgin olive oil with the area of production. The study has been conducted on almost four hundred samples of Italian extra virgin olive oil (here and henceforth, EVOO) of certified provenance.

Customers welcome safer and better food commodities, allowing higher revenues for top-quality EVOO producers; this can start a positive feedback on the production techniques which in turn reflects on the final shelf product. This moves from “low level” fraud, which can be investigated through standard analytical methods, to the more refined geographical origin hoax which, if not harmful itself for the human consumption, is discriminating factor among IGP/DOP and blend producers.⁷⁸

The isotopic data were related to climatic conditions and a geospatial model of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ distribution was developed for the authentication and verification of the geographical origin of EVOOs. This geospatial model is able to distinguish four distinct areas of EVOO production: north, south-central Tyrrhenian, central Adriatic and islands, thus highlighting a zonation of the isotopic signatures, relating the year’s growing conditions with the isotopic content. This geospatial approach can be part of a protocol for certifying EVOOs geographical origin and prevent food fraud. Limits and perspectives of the model are also discussed.

This study is published on Food Chemistry [Chiocchini et al. 2016].

5.1 EVOO Isotopic Content

Analytical methods have been developed to determine the geographical origin of food [Roßmann 2013]. While authenticity tests can work successfully with one or two stable isotope species, the determination of geographical origin involves the analysis of several stable isotopes, requiring the fulfillment of multiple conditions. A few authors have successfully exploited stable isotope techniques in the characterization, authenticity and traceability of olive oils [Angerosa et al. 1999] and more recently [Camin et al. 2010a, Camin et al. 2010b, Portarena et al 2014, Portarena et al 2015] and [Iacumin et al 2009].

⁷⁸According to Council Regulation (EC) No 510/2006 of 20 March 2006, *geographical indication* means the name of a region, a specific place that describes an agricultural product or a foodstuff and/or a preparation which takes place in the defined geographical area. See the European Community Reg. N.1151/2012, EC Reg. N.1898/2006 and EC Reg. N.510/2006.

Stable isotope traceability methodologies are based on the assumption that isotope ratios of plant integrate the features of their specific environment. The stable isotope composition of different elements in organic tissues retains important geographic and climatic information that describes the conditions of the organic matter formation. Stable isotopes thus act as ecophysiological tracers of natural processes and they are becoming increasingly used in reconstructing ecological processes and effects [West et al. 2006]. Therefore, $\delta^{13}\text{C}$, $\delta^{18}\text{O}$ and δD values in plant tissues are used to reveal their native regions.

The oxygen isotope composition of plant material reflects, along the primary assimilation pathway, both the isotopic composition of source water taken up by the roots and its ^{18}O enrichment at the leaf level, owing to transpiration through stomata. Whereas oxygen and hydrogen isotope ratios reflect the water-related processes in plants, carbon isotope ratios of plant organic matter record the environmental effects on photosynthesis [Brugnoli and Farquhar 2000], the carbon isotope composition in plants being largely determined by the photosynthetic pathway (C_3 , C_4 or CAM) fixing atmospheric CO_2 into organic matter [Lauteri et al 2004].

The general assumption in geospatial modeling for food provenance assessment is that the commodity of interest comes from one or more confined production areas, thus reflecting the peculiar isotopic composition of the provenance. These assumptions are met by some products, such as wine, honey, meat or EVOO [van der Veer 2013].

5.2 Materials and Methods

This study takes into account the isotopic composition of 387 EVOO samples from the 2009, 2010 and 2011 seasons, from Lombardia, Liguria, Toscana, Lazio, Molise, Puglia, Calabria, Sardegna and Sicilia regions, previously analysed in [Portarena et al 2014]. The actual harvesting is certified by the UNAPROL consortium of EVOO producers [UNAPROL].

Stable isotope analyses of the EVOOs were performed using an isotope ratio mass spectrometer (Isoprime, Cheadle, UK) equipped with a pyrolysis system (Euro Pyr-OH, Euro Vector Instruments & Software, Milan, Italy) and an elemental analyzer (NA1500, Carlo Erba, Milan, Italy), for measurements of $^{18}\text{O}/^{16}\text{O}$ and $^{13}\text{C}/^{12}\text{C}$ ratios, respectively.

Georeferenced raster grid layers for the isotopic composition of meteoric water were imported from [ISOMAP]. Global precipitation and temperature data (2.5 arc-minute resolution grids of minimum, average and maximum temperatures, monthly averaged, and month precipitations) were imported from [WORLDCLIM], according to the workflows in literature

[Bowen and Revenaugh 2003, Hijmans et al. 2005].

Figures 7 and 8 show the relative abundances distribution per region of carbon and oxygen, respectively.

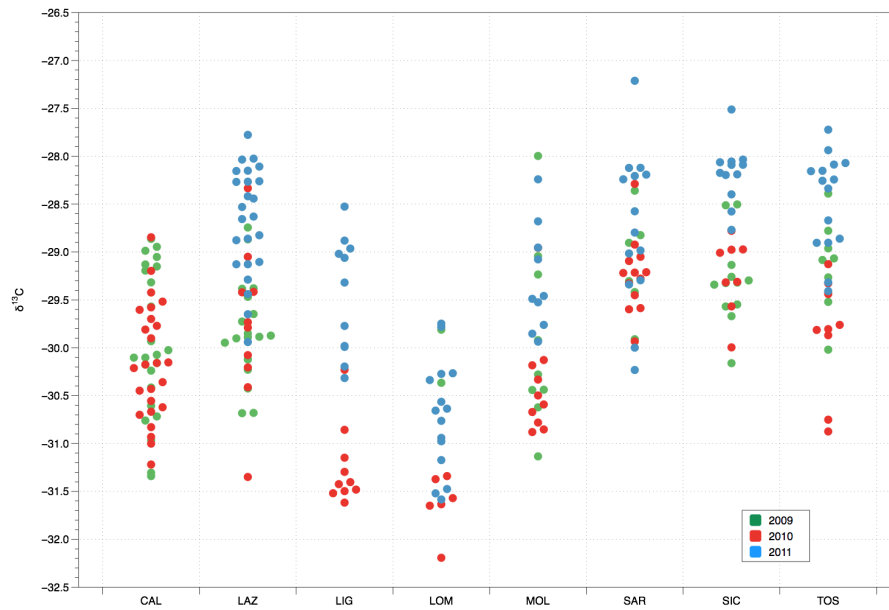


Figure 7: Measured EVOO $\delta^{13}\text{C}$ ‰ per region and year.

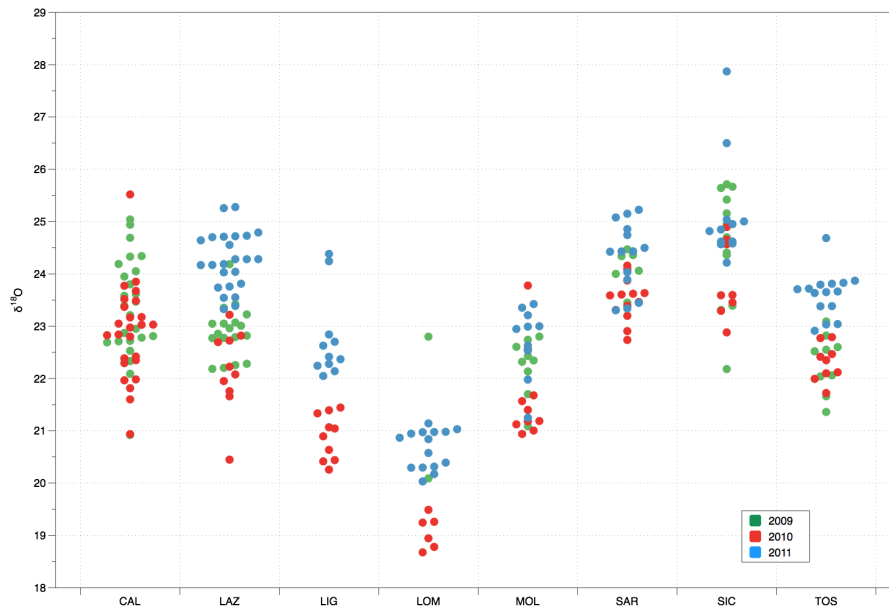


Figure 8: Measured EVOO $\delta^{18}\text{O}$ ‰ per region and year.

The xerothermic index Xi was evaluated according to the sum of the monthly Xi_{month} :

$$Xi_{month} = \begin{cases} 2T_M - P & \text{if } 2T_M - P > 0 \\ 0 & \text{otherwise} \end{cases}$$

where T_M is the average of the monthly minimum and maximum temperature and P is the

corresponding precipitation. The ^{18}O regression against precipitation and xerothermic index was calculated for each year, resulting in R^2 values from 0.4 to 0.6. (figure 9)

5.2.1 Geospatial Model

We adopted an hybrid procedure [van der Veer 2013, Bowen and Wilkinson 2002]. This procedure consists in fitting the isotope ratios distributions, evaluating the residuals with the input EVOO dataset, interpolating the residuals and summing the latter to the first interpolation. This procedure improves the confidence interval of the predicted values. The use of a Kriging interpolator allows the estimate of the model variances, too.

in detail, we followed the this procedure:

- **OLS**: Ordinary Least Squares analysis was used to quantify the spatial relationships among $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ of EVOOs for each year of production, average temperatures and pre- cipitations for the four quarters of each year, the xerothermic index and the $\delta^{18}\text{O}$ of long-term average annual precipitation, as shown in figure 9 above. $\delta^{18}\text{O} \sim Xi$ could also be better if one eliminates the zeroes of Xi from the interpolation, but we adopted a conservative approach [Cressie 1990]. The spatial autocorrelation of the regression residuals indicates the lack of a unique explanatory variable in the model, as it is almost always the case in complex ecological systems [Fortin and Dale 2011].
- **FIT**: A first estimate of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ was evaluated from the sampled EVOO values and the ancillary climatic variables. From this estimate we calculated the residuals with respect to the measured values.
- Simple Kriging of the residuals gave an interpolated grid to be used to correct the initial estimate [Oliver and Webster 2015]. It also gives a direct estimate of the model's local uncertainty.
- **SKlm**: Simple Kriging with local means [van der Veer 2013, Goovaerts 1997], added to the residuals interpolation, gave the final predicted EVOO $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ isoscapes.

In synthesis, the estimator (predicted value at a given site u of either isotope ratio) is:

$$z_{SKlm}^*(u) = f(y(u)) + \sum_{k=1}^{N(u)} \lambda_k^{SK}(u) \text{res}(u_k)$$

where $f(y(u))$ is the regression estimate at site u and the summation is the Kriging of the residuals, evaluated at u , limited to the neighbourhood $N(u)$; the $\lambda_k^{SK}(u)$ are the Kriging eigenvalues. Considering that kriging techniques directly estimate the local model uncertainty as kriging variance and the OLS tool estimates a regression variance, we assumed that the our method's variance S_{SKlm}^2 can be approximated by adding separate variance terms [van der Veer 2013]:

$$S_{SKlm}^2(u) \simeq \frac{S_R^2(u)}{N(u)^2} + S_{SK}^2(u)$$

which allows the estimate of the 95% confidence interval⁷⁹ as $1.96 \times \sqrt{S_{SKlm}^2(u)}$:

$$CI(u) = \pm 1.96 \times \sqrt{\frac{S_R^2(u)}{N(u)^2} + S_{SK}^2(u)}$$

Calculations were performed in the ArcGIS environment [ARCGIS].

5.3 Results

The range of $\delta^{18}\text{O}$ values in EVOOs was relatively large, varying from 19.1‰ to 25.1‰, over the years 2009, 2010 and 2011 [Portarena et al 2014]. A smaller range of variation, about 3.5‰, was observed for $\delta^{13}\text{C}$ (-31.6‰ to -28.2‰). The lowest mean and absolute values of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ were reported in 2010.

Samples from northern regions (Lombardia and Liguria) show the lowest values for both isotopes. EVOOs from Sicilia and Sardegna have highest values every year, while those from west-central regions, along the Tyrrhenian Sea, had intermediate values of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$. EVOOs from Adriatic regions (Molise and Puglia) were less enriched in both ^{18}O and ^{13}C compared to samples from sites located at the same latitude but along the Tyrrhenian coast.

A positive correlation was observed between $\delta^{18}\text{O}$ EVOO values of and predicted precipitation $\delta^{18}\text{O}$, with the highest correlation coefficient for 2010 ($r = 0.78$, $p < 0.01$), compared to 2011 ($r = 0.64$, $p < 0.01$) and 2009 ($r = 0.61$, $p < 0.01$) (figure 9). We observed a positive correlations between $\delta^{18}\text{O}$ values and annual mean temperature, mean temperature of the warmest months, mean precipitation of the spring quarter, respectively. The correlations were particularly robust for 2010 and 2011. A conservative, positive correlation was also found between $\delta^{18}\text{O}$ values and the xerothermic index.

⁷⁹Corresponding to a p-value 0.05. Or equivalently, 1.96 is the Z score of the 97.5 percentile of a normal distributed random variable.

No significant correlation was found between $\delta^{13}\text{C}$ values and the ancillary variables considered for the year 2009. $\delta^{13}\text{C}$ values measured in 2010 and 2011 samples were positively correlated with annual mean temperature, mean temperature of the warmest months, and the xerothermic index. In the same years, negative correlations were observed with annual mean precipitation and mean precipitation of both the spring and summer quarters.

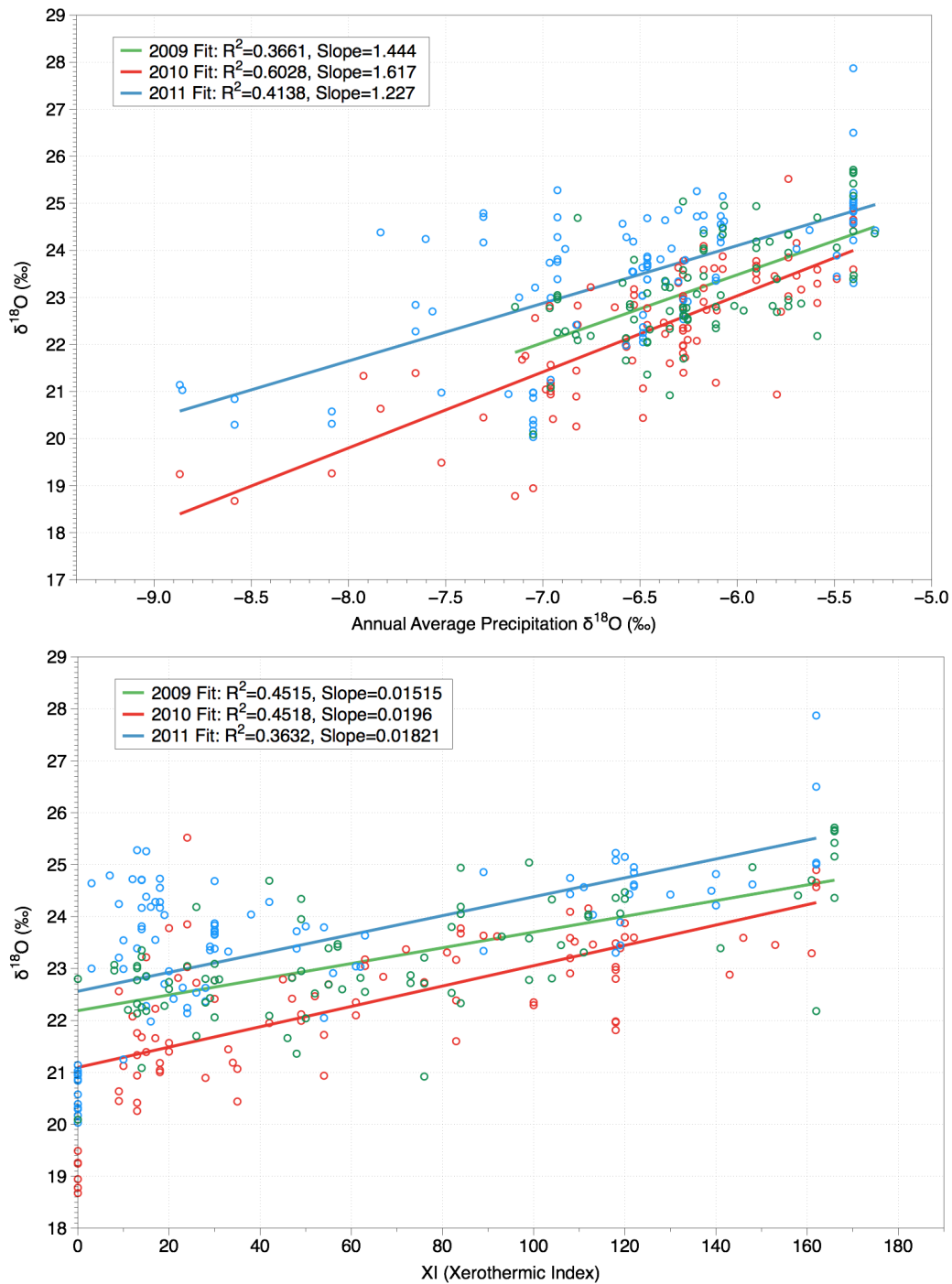


Figure 9: Measured EVOO $\delta^{18}\text{O}$ ‰ vs precipitation $\delta^{18}\text{O}$ and xerothermic index.

Since the residual variances exhibited a slight spatial autocorrelation, we incorporated

regional adjustments into the statistical model output by the geostatistical interpolation of model residuals [Bowen and Wilkinson 2002, van der Veer 2013]. The annual predicted EVOO isoscapes were composed by adding the interpolated residuals of the regression models with the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values estimated by regression. The result is shown in figure 10.

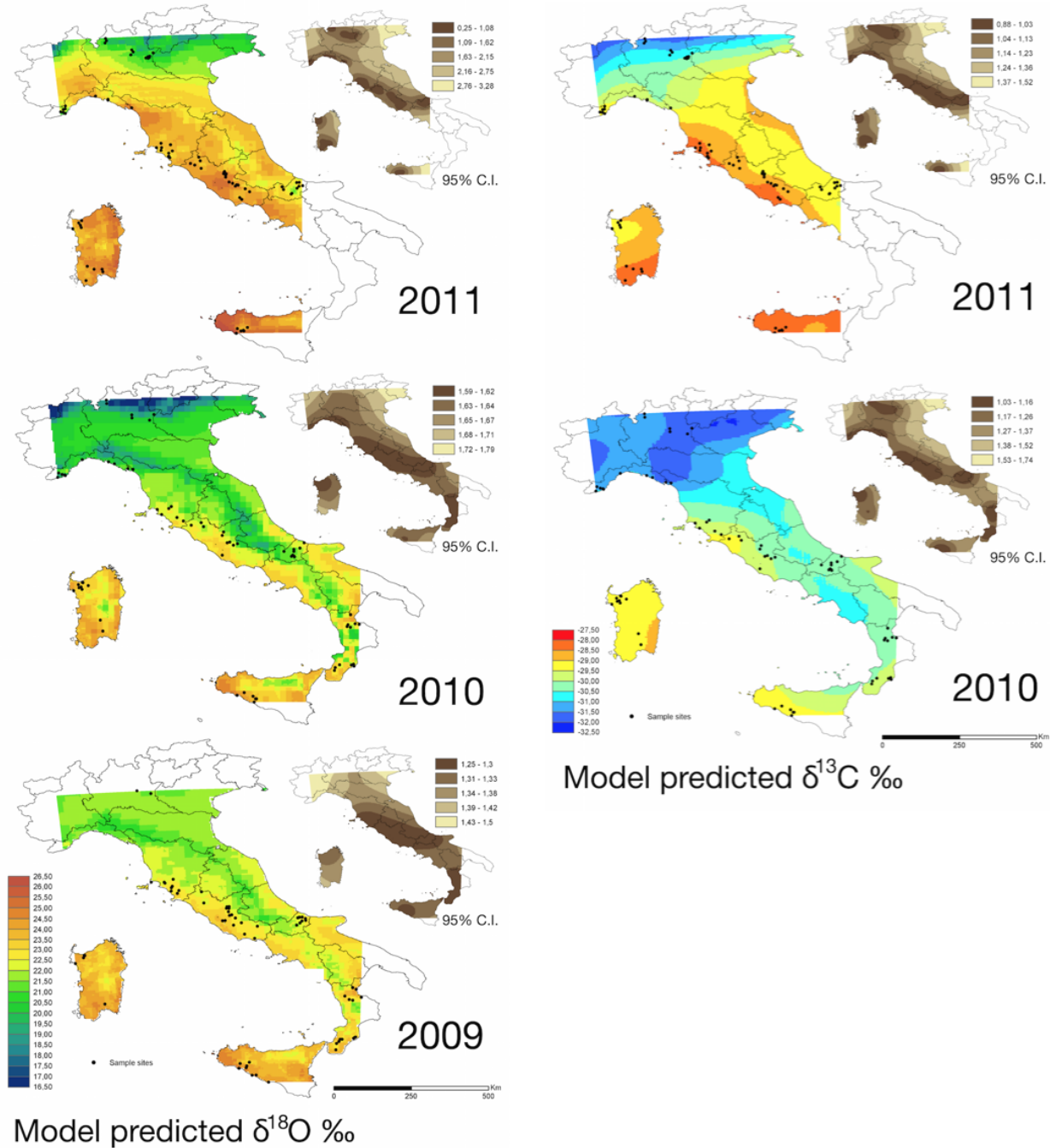
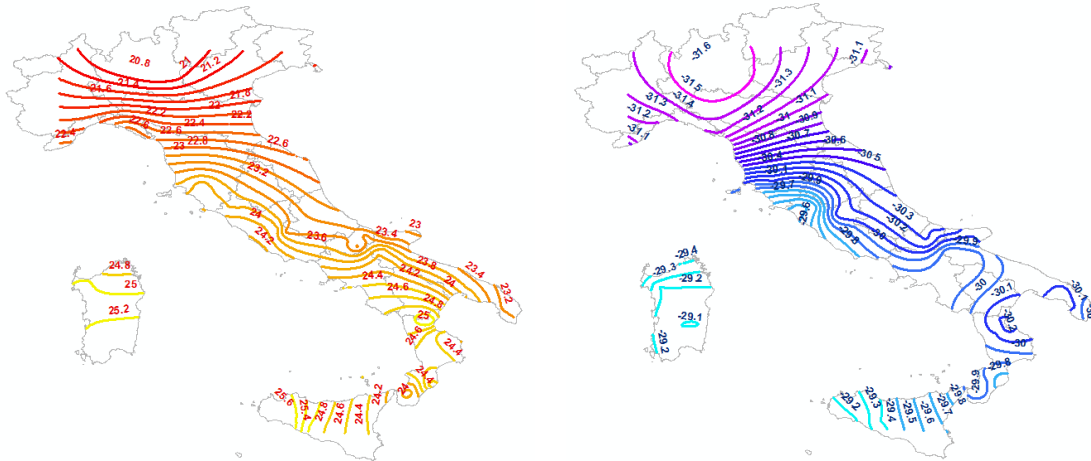


Figure 10: Predicted EVOO $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ ‰ per year. The small maps show the confidence intervals. The northeastern regions are geostatistical artefacts.

We were able to distinguish four areas of production, roughly corresponding to the Northern regions, the central Tyrrenean, Southern Adriatic and the major islands (Sardinia and Sicily). The figure below shows carbon (right) and oxygen (left) isolines for the year 2010, derived from the prediction isoscapes discussed above (figure 10, middle).



These isolines are not parallel, which indicates that $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ can be used effectively as separate statistical predictors [Cressie 1990], the most pronounced difference of pattern of the isolines is found in the southern part of Italy⁸⁰

5.3.1 Predicted ^{18}O Isoscape

The models highlight a good predictive power for the three years analysed. Although the precipitation map, based on the long-term precipitation average, does not exactly characterise the isotopic composition of plant source water, in the work we assumed it as the best isotopic composition of the average soil moisture available to olive trees, since groundwater isotope composition matches that of the long-term average of the local precipitation, although complex interactions can occur among precipitation, surface water and groundwater [Aquilina et al. 2006].

The map of the oxygen isotope composition of meteoric water in relation to the Italian peninsula, as derived from [ISOMAP] water isotopes online resource [Bowen and Wilkinson 2002], shows no evidence of a pronounced latitudinal gradient of $\delta^{18}\text{O}$ values, also the map of Longinelli and Selmo does not show any latitudinal isotopic gradient of the mean oxygen isotopic composition from Sicily to Liguria, along the Tyrrhenian coast, while they observe a slight latitudinal gradient of meteoric $\delta^{18}\text{O}$ values in Puglia (southern Adriatic coast); according to these authors, despite a few exceptions, the isotopic composition of precipitations in Italy is largely a function of elevation. [Longinelli and Selmo 2003].

Our results highlight that the geographical variability of the EVOO $\delta^{18}\text{O}$ isotopic composition mainly reflects that of the source water $\delta^{18}\text{O}$, which in turn is related to the climatic history of the production areas. The predictive model developed for $\delta^{18}\text{O}$ explains the regional component of the relationship between variables, but it is not fine enough to explain the local isotopic spatial variability. There is a positive relationship between the annual mean oxygen compo-

⁸⁰Oxygen lines are spaced 0.2‰ apart, while carbon lines are at 0.2‰, for the sake of clarity.

sition of atmospheric precipitation and the annual mean local temperature [Dansgaard 1964]. Furthermore, precipitation $\delta^{18}\text{O}$ is inversely related to the amount of rainfall [Dansgaard 1964]. Whereas the oxygen isotopic composition of plant matter depends on both the $\delta^{18}\text{O}$ of source water and enriched leaf water owing to transpiration, the local climatic conditions can be crucial in determining the relative contribution of these two drivers. The climatic conditions of the Mediterranean, characterised by seasonally low precipitation and high air temperatures, often determine partial stomatal closure in plant leaves, inducing an increase in leaf temperature. This also leads to an increase in water vapour pressure in the intercellular spaces, thus leading to a further ^{18}O enrichment of leaf water during transpiration [Barbour et al. 2005].

Despite these strong correlations, none of the climatic parameters considered in the development of the spatial model has a full predictive power. On the other hand, the xerothermic index derived from long-term averages of the climate grid layers, as well as the long-term means of $\delta^{18}\text{O}$ in precipitation show a robust predictive power for all three annual EVOO $\delta^{18}\text{O}$ models. The spatial pattern of the xerothermic index may explain the slight latitudinal gradient of the EVOOs' $\delta^{18}\text{O}$. The highest $\delta^{18}\text{O}$ values were predicted for 2011 EVOOs. In fact, the higher temperatures and the less abundant rainfall in 2011 resulted in an enrichment in both EVOO $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values, also causing high positive residuals (from +1.0‰ to +2.3‰) in e.g. central Liguria, northern Toscana and central Lazio, with respect to 2009 and 2010.

5.3.2 Predicted ^{13}C Isoscape

Unlike the oxygen, considering the 2009 samples, none of the annual climatic parameters of the model showed sufficient predictive power for $\delta^{13}\text{C}$; this is perhaps also due to the scarcity of northern samples from Liguria and Lombardia (only two samples).

As for the years 2010 and 2011, there was a clear distinction between the production of the northern regions and those from the other Italian regions. The mean monthly precipitation showed a predictive power for the $\delta^{13}\text{C}$ in 2010. The particularly abundant rainfall in 2010 could have affected the physiological response of olive trees, resulting in low values [Brugnoli and Farquhar 2000]. Monthly mean temperature is the main driver for EVOO $\delta^{13}\text{C}$ in 2011. In fact, the reduction in stomatal conductance under dry climate condition causes a decrease in the ratio of intercellular to atmospheric concentration of CO_2 , leading to an increase of $\delta^{13}\text{C}$ in photoassimilates [Brugnoli and Farquhar 2000]; as a consequence of the metabolic pathway of the synthesis of the fatty acids, these reflect the isotopic signature of the photoassimilated sugars. In fact, olive trees grown in dry environments produce EVOOs with higher

values of $\delta^{13}\text{C}$ than those produced in wetter conditions.

Despite the high spatial predictive power of the model for the EVOOS considered, the exclusion of samples from Puglia constitutes a limitation for practical applications. We hypothesise that the morphology of Puglia, the local environmental conditions, the irrigation source water and a delayed harvesting according to traditional practices could have influenced the anomalous isotopic compositions of EVOOs from this region. This highlights the need for further investigations.

5.3.3 Conclusion: Protection from *geofrauds*

Through the development of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ prediction isoscapes, we evaluated the impact of the most significant large-scale drivers for the isotopic composition of Italian EVOOs. At present, our geospatial models are able to identify EVOOs from four distinct macro-areas: north, south-central Tyrrhenian, central Adriatic and the islands (Sicilia and Sardegna).

This geospatial approach appears promising in defining a protocol for the analysis of EVOO isotopic composition, to certify their geographical origin, thus preventing food fraud. Future research should exploit a more representative sampling from Adriatic regions to better derive the isotopic composition of the EVOOs. The creation of an Italian comprehensive database of EVOO isotopic composition should be the basis for designing a comprehensive Mediterranean database of authentic samples of EVOO, to protect local producers from “geofrauds”: Using prediction isoscapes any unknown sample could be compared to a reference database for establishing the likelihood of its geographical provenance, thus ensuring consumers’ protection in accordance with the European policies on food commodities traceability.

6 Case Study: Space-time Sampling Planning

In this chapter it is described the sampling strategy of a real-life ongoing forest ecology study: the Ecuafflux project - *Análisis Integrado de los flujos de carbono en cuencas de las Andes Australes de Ecuador* (integrated analysis of carbon fluxes the southern Ecuadorian Andes catchments).⁸¹ This project deals with the carbon exchange patterns in an endangered high-mountain environment.



Figure 11: Estrellacocha catchment, Ecuadorian Andes, Elevation 4400m.

The Cajas Natinal Park, in the Southern Ecuadorian Andes (Azuay, Ecuador), hosts many little catchments. The heavy rainfall, with clouds coming both from the near Pacific Ocean (West) and the Amazonas Basin (South-East), ensures a constant abundant water supply in an otherwise desert area. It is a particularly feeble equilibrium condition which is very sensitive to the ongoing climate change.

The orogenesis is of volcanic origin, though there are no active volcanoes right now. The

⁸¹See [EQUAFLUX] web site for details.

surface morphology shows clear signs of glacial activity, with many U-shaped high mountain valleys. The catchments are located at an elevation of 3800 to 4400m. The area is crossed by the Continental Divide: most of the effluents go to the Amazonas Basin, while the western catchments waters flow to the Pacific Ocean. The only tree in the area is the *Polylepis reticulata* (*Rosaceae* family; conservation status vulnerable, according to the IUCN red list), also known as *paper tree*, due to the finely layered bark [Simpson 1979, Braun 1997].

6.1 The Project

The Ecuaflex project picked up three catchments which differ for elevation and water dissolved carbon organic compounds content for studying the delicate relationships among *P. reticulata*, shrubs and lagoons waters. Some experiments are devoted to the measurement of actual and past photosynthesis with stable isotopes techniques. In particular, they have been collected leaves, tree cores, water and soil for isotopic analyses of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$.

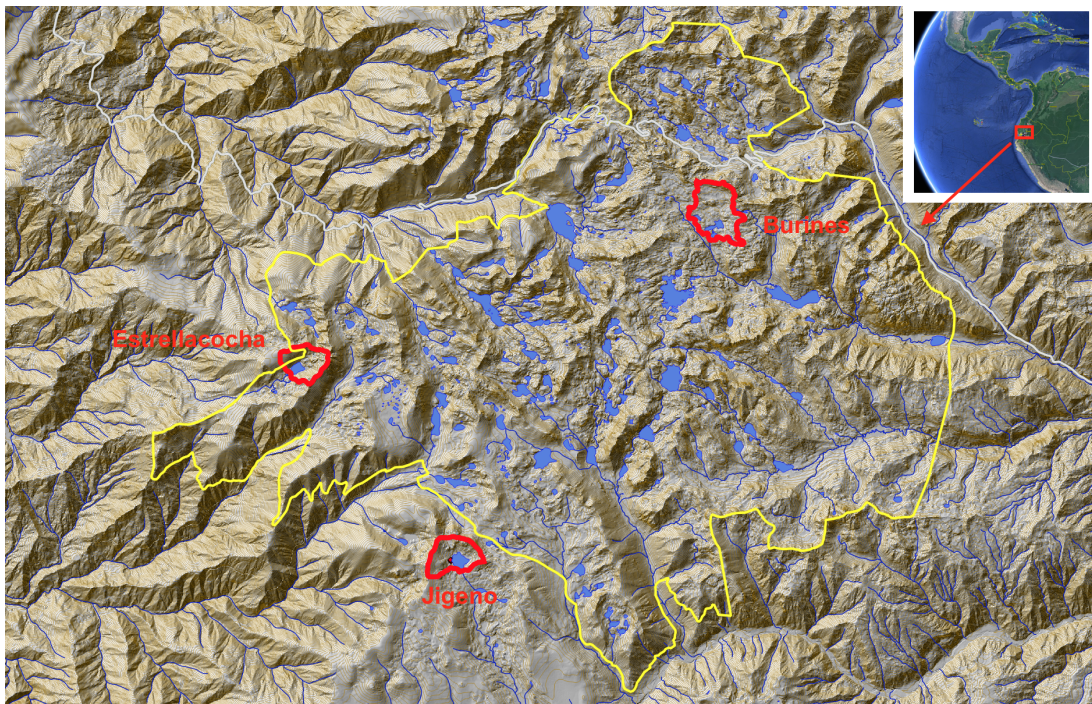


Figure 12: Cajas National Park (yellow) with the basins (red) of Jigeno and Burines, on the Atlantic side of the Continental Divide, and Estrellacocha (figure 13) on the Pacific side.

The sampling campaign started in January 2016 and is planned to last three to four years. In this project the spatial and temporal aspects of variability are interlinked and multifold. Furthermore, there are different time scales involved: the present times represented by leaves isotopic content and the tree cores which portray the historical record of photosynthesis.

6.2 The Sampling Planning and Campaign

Due to the elevation and the difficulties of access there are no known systematic studies of the area. Also the *P. reticulata* photosynthesis habits are unknown. The stable isotopes branch of the project is focused on the collection of spatially and temporally representative samples. Soil, which is very thin or sometimes totally absent, is also being sampled. This is an all-new environment that gives the opportunity to plan in advance the sampling strategy.

The Estrellacocha catchment, which is the westernmost one of the project (and of the park, too) is interested surrounded by steep slopes up to 4500m, the lake is located at 4200m and the Polylepis woods elevation ranges from the lake up to 4400m (see figure 11 at the beginning of the chapter). The temperatures measured in 2016 range from -2°C to $+18^{\circ}\text{C}$. There is no seasonality, although during the austral winter (July to September) the conditions are dryer than average. Snow is rare and there is no permanent snow coverage. To the west, the terrain slopes down to the Pacific with no other ranges in between.

Figure 13 shows the soil samples locations. The planning included a star-shaped pattern (the red dots) which was pre-planned to catch the differences among forested and bare terrain. Due to the adverse weather conditions, it was not possible to complete the sampling, anyway, the most interesting part, the woods - bare land interface, was sampled. The star pattern is important to be able to find any anisotropy in the distribution the sampled values; it also resembles a circle, which has the best area/perimeter ratio for spatial interpolations.⁸²

The sampling plan included a constant elevation and a maximum slope transect on bare ground, labelled respectively A and B in figure 13. The actual location of such transects was defined on the field. Some more points were sampled, following a maximum soil variety criterion; also these locations were defined on the field.

The original sampling plan, without previous knowledge of the terrain, contemplated a 20 - 40 - 60cm soil sampling, at least in forested areas. This was been possible, on practical grounds, since the soil is very shallow, if not completely absent. Polylepis trees grow on a rocky soil, even on boulders, and the soil structure is far from an thick, rich organic forest soil.

Sampling include also tree cores and leaves (not shown in figure 13). Bark and lichens were collected as well, as samples of *pajonal*, an ubiquitous endemic bush that grows everywhere but in polylepis forested areas.

⁸²In principle, spatial interpolations are valid only within the *convex envelope*, roughly, the perimeter, of sampled area. Out of the perimeter is extrapolation.

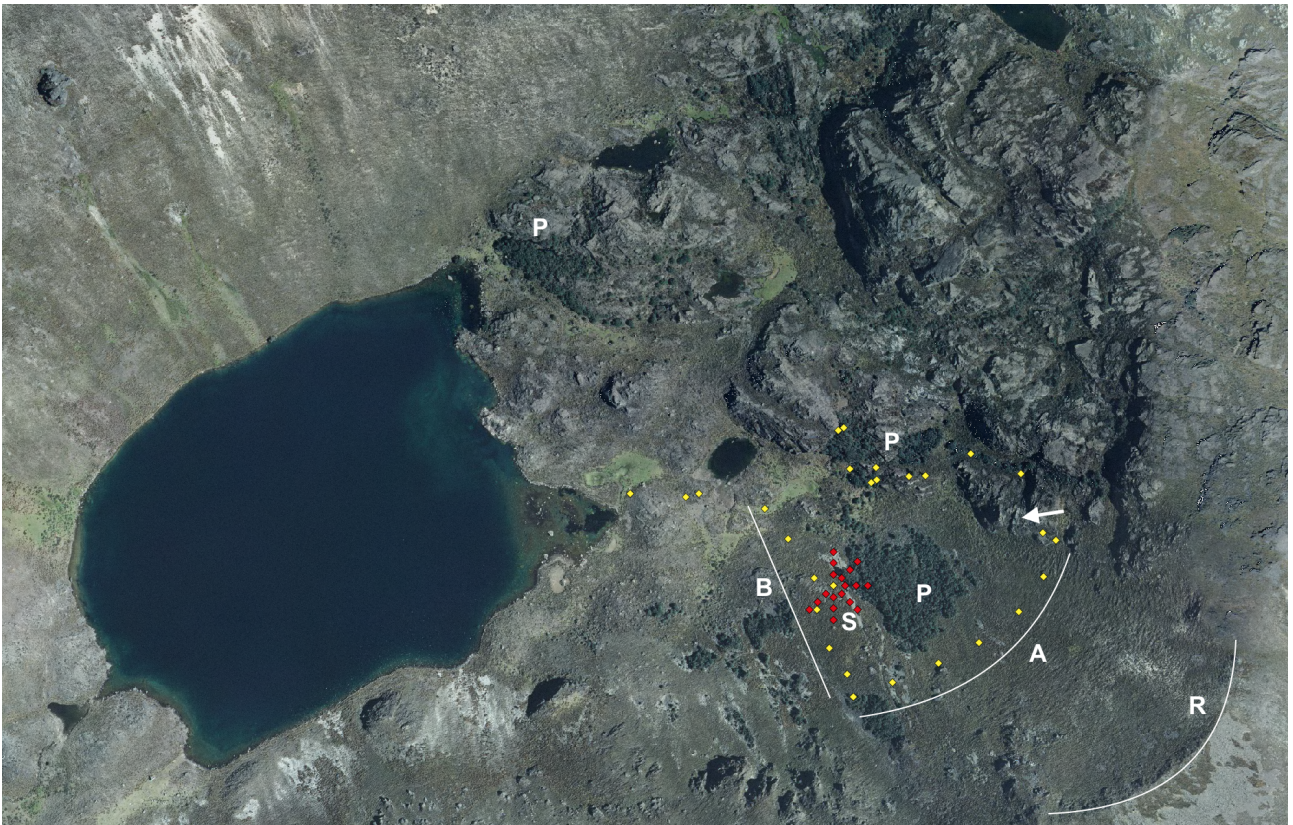


Figure 13: Estrellacocha soil sampling. P: *Polylepis reticulata* woods - S: soil isoscape - R: catchment's upper ridge (4500m elevation) - A: soil transect at 4400m - B: soil transect on the steepest line from 4400 down to 4200m. The arrow marks the shooting point of figure 11.

The spatial accuracy is about 1m, all georeferencing was performed with handheld GPS units, with exceptionally good satellite geometry.

Figure 14 shows in detail the placement of the Timescape sampling lattice. The central point is chosen as the most variable one, in terms of slope and soil conditions, preferably close to a forest interface border. The star arms are aimed at the four main directions and the intermediate ones. Four directions are not enough to catch all the possible anisotropies on a so much rugged landscape.

The most important feature to be included in the plot is a forest interface, where presumably the soil structure could show abrupt changes (a visual examination on the field reveals a darker, richer organic soil compared to the bare ground nearby).

A single campaign is enough for a single layered information (an Isoscape, or whatever else). The following campaigns, ideally two per year, will give the time development of the measurements. It is not necessary to conduct the following sampling exactly on the same sites, but a superposition of the plots of at least 50% is mandatory.

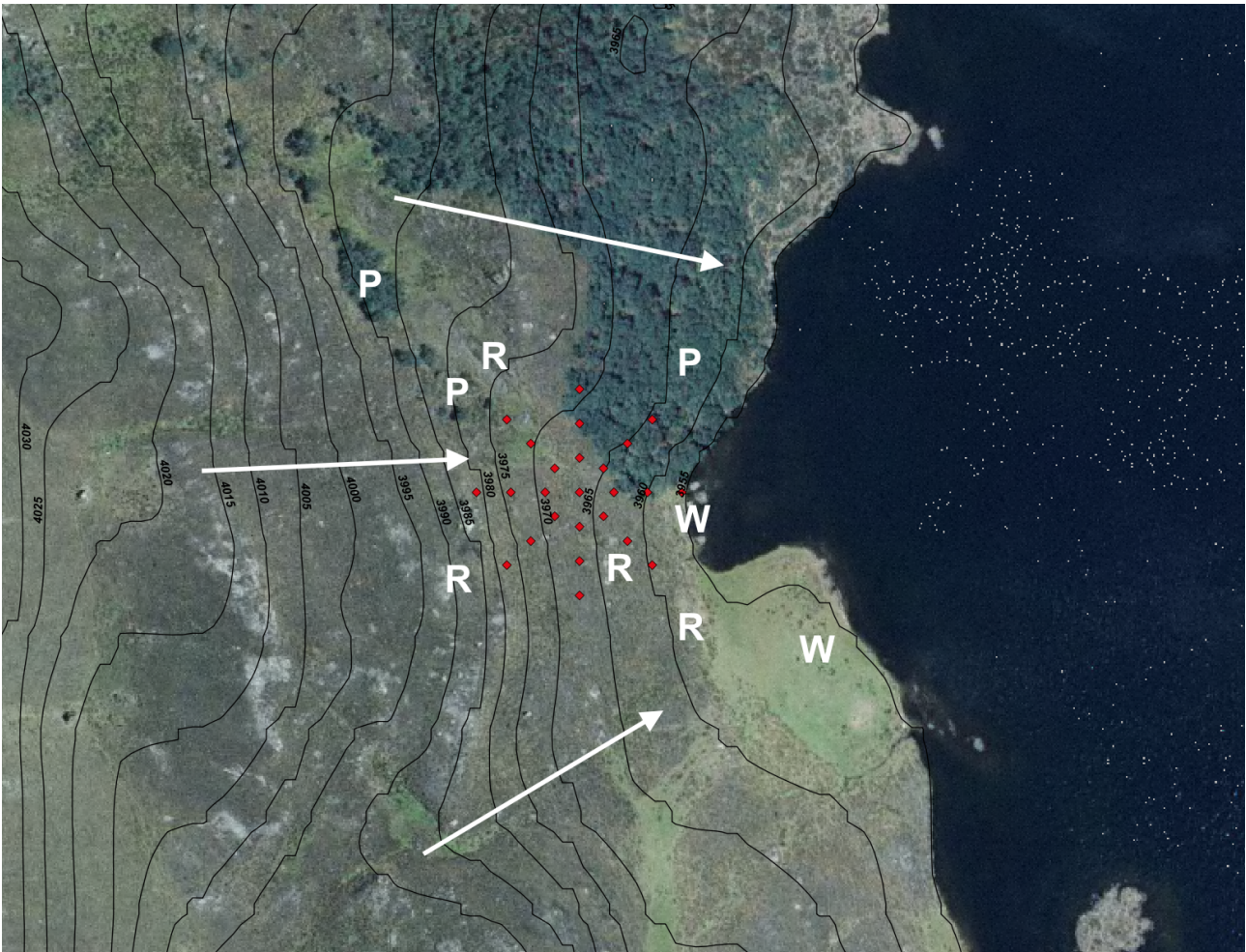


Figure 14: Jigeno soil sampling. P: *Polylepis reticulata* woods - R: bare rock areas - W: wet areas (completely soaked soil). The arrows mark the lines of maximum slope. Elevation of the plot lattice: 3955 to 3982m.

As of now only the first sampling campaign has been conducted. Other campaigns are scheduled approximately every six months. Some measurements, non involving directly stable isotopes, are performed continuously. Two dendrometers per catchment have been installed and are recording data since July 2015 (readings are conducted whenever possible).

Data collection will go on for three years, providing a unique opportunity to work with a spacetime dataset of stable isotopes measurements, complemented with phenological and environmental data. This is the ideal dataset for tuning the Timescape algorithm.

7 Conclusions and Outlook

Spatial and temporal patterns in ecological systems are among the most complex phenomena to be modelled. Statistically, countless disturbance sources worsen an already noisy situation, raising the need of some ad hoc procedures.

In the last decade, the widespread availability of cheap computing power opened up the path to the implementation of once-specialized geostatistical algorithms in the field of forest ecology. The evolution of sensors and recording stations, sometimes equipped with an on-board GPS, brought into play the need of time series analysis, too.

A feature of modeling in the field of ecological systems is the simultaneous presence of space and time variability. While established tools are available for both spatial modeling and time series analysis, there is a lack of simple instruments whenever both the sources of variability have to be considered on equal grounds.

Sampling in a natural environment also adds a peculiar *flavour* to the observations dataset: the records are often sparse, both in space and time, due to the nature of the sampling which, unlike a lab-based experiment, is subject to many disturbances. Instead of trying to achieve an unpractical precision during the sampling, it is better to be able to treat a dataset that has been collected in less-than-ideal conditions.

A number of consolidated techniques can be used efficiently for both space and time modeling alone. Sometimes one is able to pick up a single (better, a predominant) source of variability, neglecting the other. This is by far the most common procedure, which is often correct.⁸³ Most times one builds a series of independent spatial models taken at selected times, like the frames of a movie.

Less common, but conceptually equivalent, is the time variability analysis of observation that have been grouped following a spatial average criterion, which is often questionable.

Sometimes it is not possible to privilege a single source of variability. One needs to take into account space and time on the same grounds. However, there are no simple solutions here.

The idea of the **Timescape Algorithm** is to provide a reasonably simple tool for the treatment of such spacetime variability. This tool has been developed, borrowing from the field of statistical physics, with some constraints:

Ease of Use: it can be run on most computers.

⁸³See the Extra Virgin Olive Oil case study, chapter 5, where each year of has been treated on its own.

Ecology-Oriented: it is focused on the needs of ecological modeling.⁸⁴

Simplicity: it resembles an ordinary geostatistical interpolator.⁸⁵

Affordability: the software is distributed with an open license (GNU-GPL v3.0).

Reliability all the software's code is available and expandable.

GIS Compatibility: The software has been designed to be the slightest possible detour from a consolidated GIS workflow, since more and more researchers use a GIS environment for both data storage and modeling.

The Timescape Algorithm is a possible response to a wide class of problems in ecological modeling, those which involve spatiotemporal patterns of change. It is not by far the answer to all the modeling issues of forest ecology.

Almost all data sampling is conducted nowadays with the aid of a GPS, thus providing a space- and time-labelling of the samples, so that it is not hard to imagine the future availability of more and more candidates for modeling, also from studies that are focused on other ecological topics. So presumably the need for tools like Timescape will increase in the next few years. The open source distribution of the software allows any researcher to include Timescape interpolations in his/her study.

A short period outlook includes the release of a new revision of **Timescape Local**, the projected coordinates version of the software, and extensive modeling of stable isotopes in precipitation, from the GNIP archive [GNIP]. The kind of coverage offered by the GNIP precipitation data is the ideal playground for testing the capabilities (and the limits) of the Timescape approach at a global scale. Some Mediterranean-sized models will be evaluated to investigate the character of precipitation waters' isotopes on Italy, which is notoriously complex, due to the peculiar orography of the country.

As of now, the Timescape Algorithm is an estimator of a single variable,⁸⁶ it is possible, as a matter of principle, to extend it in a multivariate fashion; however, the added computational and methodological complexity suggest a comprehensive testing of the one-variable version before moving on.

⁸⁴i.e. complexity vs precision to the last digit.

⁸⁵A *minimum* added complexity is due to the presence of the causal constraints. See the causal cone structure, chapter 2, and appendix A for the mathematical details.

⁸⁶One can consider any number of non-interacting variables (no covariance) Ancillary variables can be used as independent variables, too.

A Appendix: The Timescape Maths

This rather technical appendix describes in detail all the maths underlying the Timescape Algorithm and its possible implementations. Part of the appendix is referred to the **Timescape-Global** published software package. The actual data structure is described in Appendix B.

A.1 Space and Time Distances

The question of what should be considered a measure of distance in a spacetime is not a trivial one. The Minkowskian spacetime of Physics is centred on the invariance of the speed of light and it is not suitable for the kind of problems encountered in forest ecology; we will introduce a set Euclidean measures which suite our needs. There is nothing “relativistic” in Timescape, although its construction resembles (and in fact it is borrowed from) the Minkowskian double cones. In fact, we will not define the distance, but rather a set of suitable distances.

Following the current Minkowskian terminology, we call an **event** $x = (t, \mathbf{x})$ any point of spacetime X , where t is the time coordinate and $\mathbf{x}$ are the spatial coordinates. X has not a definite topology. It can be simply $\mathbb{R}_0^+ \times \mathbb{R}^2$ (flat space) or something topologically equivalent to $\mathbb{R}_0^+ \times \mathbb{S}^2$ (a sphere, an ellipsoid or something more general, too).⁸⁷

Starting from the time, the distance d_t is very simple to define. Given two events $x = (t_x, \mathbf{x})$ and $y = (t_y, \mathbf{y})$, the distance inherited from $\mathbb{R}$ could be defined as $|t_x - t_y|$ which can be made homogeneous with the space components by multiplying it for a parameter c having the dimensions of a speed: $c|t_x - t_y|$. Or, more generally, we allow c to be a function of the events:

$$d_t(x, y) = c(x, y) |t_x - t_y| \quad (5)$$

The function c in (5) should be constrained in order to obey all the constraints for a distance function (6). Every positive constant will do, as some monotonously increasing functions can do, but not all of them (the triangle inequality could fail to be true). Constraints-violating functions can be used, but this “breaks the rules” somehow. A remarkable exception is the use of harmonic functions in c : these are obviously not well-behaved distances but can be of great help in modeling periodic phenomena.

The spatial part of the distance $d_s(x, y)$ is more complicated to define. It depends on the topology of the space, so the use of a Riemannian metric is due, in principle. The point is that

⁸⁷ $\mathbb{R}$ is the set of real numbers, $\mathbb{R}_0^+$ is the set of positive real numbers, including the zero value and $\mathbb{S}^2$ is the two-dimansional sphere. With *sphere* here we mean the layman’s surface of the sphere only.

evaluating the geodesic distance between two events is time-consuming and in most of the cases not worth the effort. As a general rule, a distance should satisfy the following constraints:

$$\begin{aligned}
 d(x, y) &\geq 0 \quad \forall x, y \in X && \text{non-negativity} \\
 d(x, y) &= d(y, x) \quad \forall x, y \in X && \text{symmetry} \\
 d(x, y) &= 0 \quad \text{iff } x = y && \text{coincidence} \\
 d(x, z) &\leq d(x, y) + d(y, z) \quad \forall x, y, z \in X && \text{subadditivity}
 \end{aligned} \tag{6}$$

The formal solution consists in measuring the geodesic line from x to y (or vice versa, the result does not change for symmetry). This is achieved solving the equation for the path $x(\tau)$ parametrised by τ ⁸⁸ (sums over repeated indices understood):

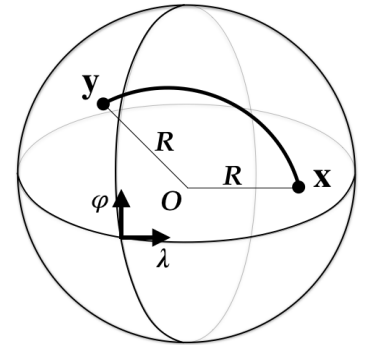
$$\ddot{x}^\lambda + \Gamma_{\mu\nu}^\lambda \dot{x}^\mu \dot{x}^\nu = 0 \tag{7}$$

where $\dot{x}^\lambda$ and $\ddot{x}^\lambda$ indicate the first and second derivatives of the λ component of x relative to τ and $\Gamma_{\mu\nu}^\lambda$ is the Christoffel symbol which is related to the space metric $g_{\mu\nu}$ through

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\rho} \left(\frac{\partial g_{\rho\mu}}{\partial x^\nu} + \frac{\partial g_{\rho\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\rho} \right)$$

Numerically solving (7), though feasible in principle, is not compatible with acceptable running times, at least not with a standard desktop hardware.⁸⁹ Compromises are in order, here. The shape of the Earth can be approximated with an oblate ellipsoid, for which exact solutions exist, but they involve the evaluation of elliptic integrals; these can be evaluated numerically, but it is still too much for an ordinary computer in a reasonable time.

Another little step further is the approximation of the shape of the Earth with a sphere of radius R . In this case a simple solution exists to evaluate the length of the shortest arc connecting two points $\mathbf{x} = (\lambda_x, \varphi_x)$ and $\mathbf{y} = (\lambda_y, \varphi_y)$ lying on the surface, it is the length of the arc of circle (the geodesic line on a sphere) connecting $\mathbf{x}$ and $\mathbf{y}$:



$$d_s(x, y) = R \arccos \left(\sin \varphi_x \sin \varphi_y + \cos \varphi_x \cos \varphi_y \cos(\lambda_x - \lambda_y) \right)$$

⁸⁸A sort of proper time, following the lines of relativistic lexicon.

⁸⁹The calculation has to be repeated a huge number of times: the number of space cells of the model times the number of source points: something easily of the order of magnitude of billions.

where λ is the longitude and φ the latitude of the points. This is in fact the Earth radius R times the aperture $\widehat{\mathbf{xOy}}$ of the arc $\widehat{\mathbf{xy}}$. This is easy to evaluate since involves only five trigonometric functions and an inverse one; there is no numerical integration (oblate ellipsoid) or solution of differential equations (general case) involved. Curved surfaces coordinates are often singular in one or more points, this is in fact the case also with spherical coordinates, where the poles are singular. In fact, we have a periodic (the longitude λ) coordinate and two singular points. This is relatively harmless but it introduces some *caveat* for the evaluation:

- Periodic coordinates should be checked before evaluation: they should be reduced to the fundamental interval (say $[0, 2\pi)$ or $(-\pi, \pi]$ for the longitude). Sometimes the trigonometrical functions do the trick automatically, but it is always better to act prudentially.
- Singular points often lead to oversampling, the meridians shrinking about the poles are typical. This is not bad *per se* but it can be time-consuming.
- Flattening, i.e. projecting coordinates should always be taken seriously into account, according to the nature of the surface involved. If the phenomenon under investigation is not worldwide or at least spread over a continent-sized area it is always better to use projected coordinates (UTM, Lambert conical, etc).

Now applying the Pythagorean theorem to the two distances we obtain

$$d(x, y) = \sqrt{d_t^2(x, y) + d_s^2(x, y)} \quad (8)$$

which in spherical coordinates reads

$$d(x, y) = R \sqrt{\left(\frac{c}{R}\right)^2 (t_x - t_y)^2 + \arccos^2 \left(\sin \varphi_x \sin \varphi_y + \cos \varphi_x \cos \varphi_y \cos(\lambda_x - \lambda_y) \right)} \quad (9)$$

when c is a constant. The c/R ratio encodes a fundamental information,⁹⁰ about the reciprocal roles of space and time variability, otherwise said, large c/R ratios privilege the time component of variability, while small ratios depress the role of the time variability with respect to space.

The **TimescapeGlobal** software employs (9) for the evaluation of spacetime distances between events. The use of projected coordinates as in **TimescapeLocal** is straightforward in that we can use a simple Euclidean sum.⁹¹

⁹⁰The c/R ratio has dimensions $[LT^{-1}]/[L] = [T^{-1}]$, like a frequency. It is roughly related the time needed for a phenomenon to spread worldwide.

⁹¹**TimescapeLocal** is not yet released, a development α version is available upon request to the author.

If we have a flat N -dimensional space the distance can be the simple Euclidean one:

$$d(x, y) = \sqrt{\sum_{k=1}^N (s_x^k - s_y^k)^2 + c^2(t_x - t_y)^2} \quad (10)$$

where s^k is the k -th space component of a point. Aside from a factor c , this is the ordinary Euclidean N -dimensional distance, not to be confused with the Minkowskian relativistic distance, which is different for having a sign switched between the spatial and temporal parts.

It is worth mentioning that more distances can be used other than the Euclidean one. In flat space we can use the equivalent diamond metric

$$d(x, y) = \sum_{k=1}^N |s_x^k - s_y^k| + c |t_x - t_y|$$

or the square metric

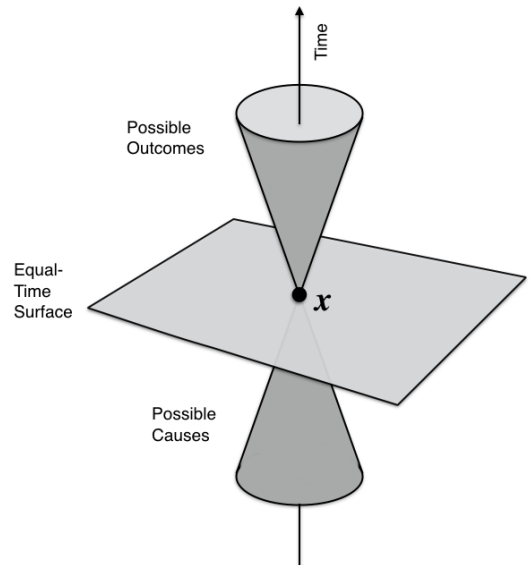
$$d(x, y) = \max \left\{ \max_k \left\{ |s_x^k - s_y^k| \right\}, c |t_x - t_y| \right\}$$

or a fancier mixing of these, combining a Euclidean sum of d_t with the diamond/square version of d_s . All these metrics are equivalent to the Euclidean one.

A.2 Causal Structure

The key feature of the Timescape Algorithm is *causality*. The spacetime structure described in the previous section is, aside from a factor c , just ordinary three-dimensional space. The time still needs to be singled out as the “direction” followed by the patterns of change. We have to plug a causal structure by hand into the spacetime in order to drive the change towards a *forward only* direction. The Minkowskian metric⁹²

$$d_M^2(x, y) = c^2(t_x - t_y)^2 - \|\mathbf{x} - \mathbf{y}\|^2$$



⁹²That, strictly speaking, is not a metric since it is not non-negative.

has a natural interpretation in terms of causality: a positive $d_M^2(x, y)$ means that y falls inside the *causal double cone* of x , either in its past ($t_y < t_x$) or in its future ($t_y > t_x$), while a negative $d_M^2(x, y)$ means that there is no causal connection between y and x .

Our Euclidean spacetime is topologically different from a Minkowskian one, so the causal double cone structure cannot emerge in a natural way. We must impose the causal structure somewhat artificially. Let define a *cone aperture* k which, in principle, can be a positive function of the events $k(x, y)$ corresponding to the maximum acceptable ratio between the space and time components of the distance:

- y can be a *cause* of x if $d_s(x, y) \leq k d_t(x, y)$ and $t_y < t_x$
- y can be an *outcome* of x if $d_s(x, y) \leq k d_t(x, y)$ and $t_y > t_x$
- y has no causal connection with x if $d_s(x, y) > k d_t(x, y)$

The aperture of the cone, or of the cone-like structure if $k(x, y)$ is not a constant, defines the strictness of the causality constraints: the broader the cone (a big k) the looser the constraints; the narrower the cone (a small k) the stricter the constraints. This structure, unlike the rigid Minkowskian one, allows for the definition of a flexible concept of causality.

An infinite value of k corresponds to a double cone spanning all the spacetime, so that any event can be in principle connected with all the others. On the other hand, $k = 0$ means that there is no spread of causality in space, a complete static solution where the influence is limited to the *time line* passing through x . Allowing $k = k(x, y)$ means that the causal constraints can vary from place to place and over time.

Now we must find a way to translate mathematically the constraints mentioned above in an easily computable way. The key is the Heaviside theta or step function:⁹³

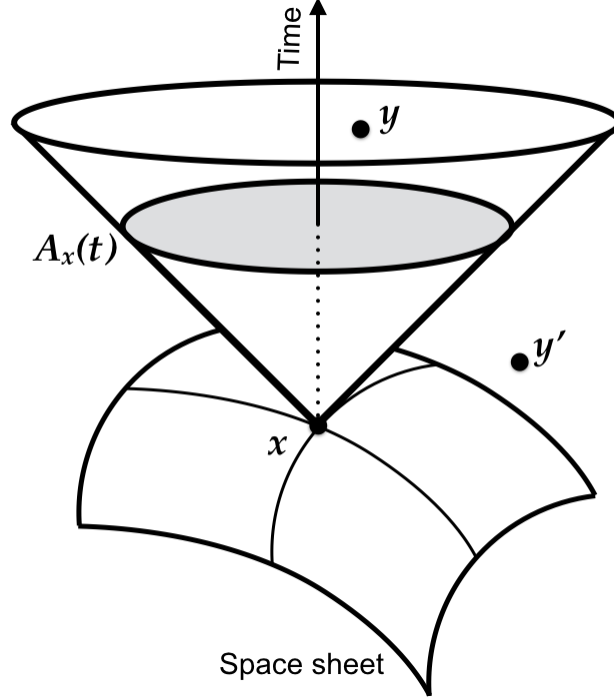
$$\theta(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

this function acts as an on/off switch so that

- y can be a *cause* of x if $\theta(k d_t(x, y) - d_s(x, y)) \theta(t_x - t_y) = 1$
- y can be an *outcome* of x if $\theta(k d_t(x, y) - d_s(x, y)) \theta(t_y - t_x) = 1$
- y has no causal connection with x if $\theta(k d_t(x, y) - d_s(x, y)) = 0$

⁹³There is not consensus in the literature about the value of $\theta(0)$.

A straightforward interpretation is that $k d_t(x, y)$ represent the maximum possible *area of influence* $A_x(t)$ of the event x at a given time t , while $d_s(x, y)$ is the actual spatial separation between such events. Whenever y falls into this $A_x(t)$, the contribution of x has to be taken into account, while when an event y' falls out of $A_x(t)$ (so out of the forward causal cone of x), the contribution of x is null.



For any t the set $A_x(t)$ represent a *slice* of K_x^+ , the *forward causal cone* of the event x , which can be seen as an infinite union

$$K_x^+ = \bigcup_{\tau \in \mathbb{R}_0^+} A_x(t_x + \tau) \quad (11)$$

Equivalently, the *backwards causal cone* of the event x is

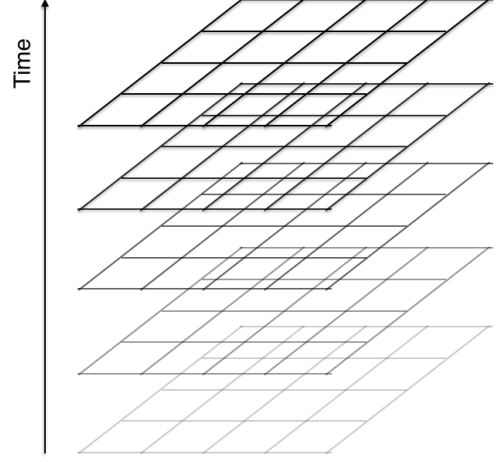
$$K_x^- = \bigcup_{\tau \in \mathbb{R}_0^+} A_x(t_x - \tau) \quad (12)$$

In summary K_x^+ and K_x^- , defined by (11) and (12), are the mathematical representation of the possible outcomes and the possible causes of the event x . To compute a Timescape model we can follow two strategies: the first consists in picking one source event at a time from the samples dataset and try to find what happens in its K_x^+ . The second strategy consists in subdividing the model in a set of discrete elements (cells of spacetime) centred about an event y and find which source points form the samples fall into K_y^- ; this is the computationally recommended one.

A.3 Building the Model

A Timescape model is a collection of voxels⁹⁴ equipped with a value. The extent of the model is the product of a time interval $[T_m, T_M]$ times the spatial extent of the surface involved. In the case of a longitude/latitude parametrisation of the surface of the Earth the extent of M is $[T_m, T_M] \times [\Lambda_m, \Lambda_M] \times [\Phi_m, \Phi_M]$, where Λ and Φ stand for the limits in longitude and latitude.

Each *time sheet* of voxels corresponds to a spatial replica at a given time.⁹⁵ If M is composed by $N_T \times N_\Lambda \times N_\Phi$ elements, the n th time sheet has time



$$t_k = T_m + \frac{T_M - T_m}{N_T} \left(k + \frac{1}{2} \right), \quad k = 0 \dots N_T - 1 \quad (13)$$

with the $\frac{1}{2}$ bias correction factor to make the sheets match the centres of the corresponding plane of voxels. Following the same lines of surgery, every sheet is subdivided in pixels of centres ⁹⁶ (λ_i, φ_j) of coordinates

$$\begin{aligned} \lambda_i &= \Lambda_m + \frac{\Lambda_M - \Lambda_m}{N_\Lambda} \left(i + \frac{1}{2} \right), \quad i = 0 \dots N_\Lambda - 1 \\ \varphi_j &= \Phi_m + \frac{\Phi_M - \Phi_m}{N_\Phi} \left(j + \frac{1}{2} \right), \quad j = 0 \dots N_\Phi - 1 \end{aligned} \quad (14)$$

A discrete Timescape model M consists of a finite set of voxel-value pairs m_{kij}

$$M = \{m_{kij}\} = \left\{ (y_{kij}, v_{kij}) \mid y_{kij} \in \mathbf{ST}, v_{kij} \in \{null\} \cup \mathbb{R} \right\}$$

where y_{kij} is an event of the spacetime $\mathbf{ST}$ and v_{kij} is its value. The value is calculated from the samples following some rules and can be *null* if none of the samples falls within $K_{y_{kij}}^-$, supposing that the modelled phenomenon can be represented with real numbers.

The *samples collection* is the set $S = \left\{ (x_n, v_n, \mathbf{a}_n) \mid x_n \in \mathbf{ST}, v_n \in \mathbb{R} \right\}$ of samples at event x_n , with a value v_n and possibly an array $\mathbf{a}_n$ of associated ancillary variables.⁹⁷

Each m_{kij} is evaluated independently of all other elements of M , so that a model can be

⁹⁴A voxel is a discrete three-dimensional *volume element* as a pixel is a discrete two-dimensional *picture element*.

⁹⁵Unlike the Minkowskian case, here the time is absolute.

⁹⁶We use λ and φ to label the space components, not necessarily addressed by longitude and latitude.

⁹⁷The v_n must be defined but the ancillary variables can be *null*.

evaluated sheet-by-sheet, as is the case of **TimescapeLocal** and **TimescapeGlobal** or in any other manner. Parallelisation is straightforward.

The *spacetime support* $\mathbf{S}$ of S is the finite set $\{x_n\}$ of the sample points events. In the following, with a slight abuse of notation, we will use S for both the samples collection and its spacetime support.

For any y_{kij} (y for short) we can interpolate a value, provided that there are sample events into its backwards cone: $K_y^- \cap S \neq \emptyset \Rightarrow v_{kij} \neq null$. In fact, defining the weight function⁹⁸ $w(d; y, x, v, \mathbf{a})$ and the estimate function $f(v; d, y, x, \mathbf{a})$ we have

$$v_{kij} = \left[\sum_{K_y^- \cap S} w(d; x_n, y_{kij}, v_n, \mathbf{a}_n) \right]^{-1} \sum_{K_y^- \cap S} w(d; x_n, y_{kij}, v_n, \mathbf{a}_n) f(v_n; d, y_{kij}, x_n, \mathbf{a}_n) \quad (15)$$

where $d = d(x_n, y_{kij})$, of course. The normalised summation in (15) resembles a trivial inverse distance weighted estimation (IDW); this would be true if we choose the trivial $w = 1/d$ as the weight estimator, but this limitation can be easily overcome, by defining suitable weight functions. The summations in (15) are extended to the backwards causal cone of y_{kij} .

In the simplest case of the standard IDW interpolation, when the weight is given by $1/d^\alpha$, where α is a positive constant, (15) becomes

$$v_{kij} = \frac{\sum_S \theta(kd_t - d_s) \left(\sqrt{d_t^2 + d_s^2} \right)^{-\alpha} f(v_n; \sqrt{d_t^2 + d_s^2}, y_{kij}, x_n, \mathbf{a}_n)}{\sum_S \theta(kd_t - d_s) \left(\sqrt{d_t^2 + d_s^2} \right)^{-\alpha}}$$

which if $\alpha = 1$, the estimate function f depends only on v and d_t is simply $c\Delta t$ reduces to

$$v_{kij} = \frac{\sum_S \frac{\theta(c k \Delta t - d_s)}{\sqrt{c^2 \Delta t^2 + d_s^2}} f(v_n)}{\sum_S \frac{\theta(c k \Delta t - d_s)}{\sqrt{c^2 \Delta t^2 + d_s^2}}} \quad (16)$$

that represents an almost trivial IDW interpolation with a causal constraint added through the θ functions. in the limit $k \rightarrow \infty$ equation (16) becomes an ordinary IDW in three-dimensional space, the causal constraint reducing simply to considering only the past sample events from the samples dataset.⁹⁹

⁹⁸The w depends basically on the spacetime distance of the events, and it can also be a function of the events positions (thus accounting for inhomogeneities and anisotropies) and the samples' value and ancillary values too. The f depends basically on the value of the sample (trivially the value itself) but can depend also on the other quantities.

⁹⁹Also this last constraint can be released, allowing back-and-forth action in time.

A.4 Evaluation Algorithm

In order to build a model M we start from a the finite samples collection S which represent the actual observations (i.e. the physical measurements) of the investigated phenomenon and proceed as follows:

The first step consists in the definition of set of coordinates $\{(t_k, \lambda_i, \varphi_j)\}$ representing the events located at the centres of the discrete events which constitute M , given a *time sheet size* w_s and a *space cell size* w_c . These coordinates are grouped in the three arrays $\underline{t}$, $\underline{\lambda}$ and $\underline{\varphi}$:

Algorithm 1 Preliminary steps

```

procedure COORDINATES( $w_c, w_s, T_m, T_M, \Lambda_m, \Lambda_M, \Phi_m, \Phi_M$ )
   $\delta t \leftarrow w_s$ 
   $t_0 \leftarrow T_m + \delta t/2$ 
  while  $t_n \leq T_M$  do ▷ voxel sheets times  $\underline{t}$ 
     $t_{n+1} \leftarrow t_n + \delta t$ 
   $\delta \lambda \leftarrow w_c$ 
   $\lambda_0 \leftarrow \Lambda_m + \delta \lambda/2$ 
  while  $\lambda_n \leq \Lambda_M$  do ▷ voxel space coordinates  $\underline{\lambda}$ 
     $\lambda_{n+1} \leftarrow \lambda_n + \delta \lambda$ 
   $\delta \varphi \leftarrow w_c$ 
   $\varphi_0 \leftarrow \Phi_m + \delta \varphi/2$ 
  while  $\varphi_n \leq \Phi_M$  do ▷ voxel space coordinates  $\underline{\varphi}$ 
     $\varphi_{n+1} \leftarrow \varphi_n + \delta \varphi$ 
return  $\{\underline{t}, \underline{\lambda}, \underline{\varphi}\}$ 

```

Then the actual model is evaluated, one voxel at a time.¹⁰⁰ As algorithm control “tuning knobs” we add three further parameters:

- A space distance threshold D_s so to discard from the summations all the events too spaced apart; this is like switching off any interaction for events farther than D_s .
- A maximum number of allowed near primes N_{max} : for any y_{kij} of M we use at most N_{max} elements of S for the calculations.
- A *picking probability* $\mathcal{P}$ which introduces some randomisation spicing to an otherwise deterministic procedure. Any $x_n \in K_{y_{kij}}^-$ is included in the summation only with probability¹⁰¹ $\mathcal{P}$ to allow the users check the stability of their datasets: The algorithm can be

¹⁰⁰Operatively, all voxels are inserted empty in the database before their actual evaluation. This is done in order to minimize the database file growing / shrinking during the run. Voxels records will just be updated with the appropriate values after evaluation: this does not change the record size.

¹⁰¹A random number in $[0, 1]$ is generated at each step of the summation loop and it is checked against $\mathcal{P}$.

run many times and if the dataset is said to be *stable to* $(1 - \mathcal{P})$ with precision σ if the variances of all the estimated elements of M are less than σ .

This said, the general model loop goes as follows:

Algorithm 2 Timescape Model

```

procedure MODEL( $\underline{t}, \underline{\lambda}, \underline{\varphi}, \mathcal{E}, N_{max}, \mathcal{P}, D_s, c, k, w, v$ )
  for  $t_k \in \underline{t}$  do
    for  $\lambda_i \in \underline{\lambda}$  do
      for  $\varphi_j \in \underline{\varphi}$  do
         $y_{kij} \leftarrow (t_k, \lambda_i, \varphi_j)$  ▷  $y$  for short
         $K_y^- \leftarrow \emptyset$  ▷ find  $K_y^-$ 
        for  $x_n \in S$  do
          if  $\text{RAND} \leq \mathcal{P}$  then ▷ random picking, always true if  $\mathcal{P} = 1$ 
             $d_t \leftarrow c(x_n, y) |t_{x_n} - t_y|$ 
             $d_s \leftarrow \int_{\Gamma} \sqrt{g} d\tau$  ▷ geodesic length
            if  $d_s \leq \min\{k(x_n, y) d_t, D_s\}$  then
               $d_n \leftarrow \sqrt{d_t^2 + d_s^2}$ 
               $K_y^- \leftarrow K_y^- \cup \{(x_n, d_n)\}$  ▷ one more causal event
          if  $K_y^- \neq \emptyset$  then
             $K_y^- \leftarrow \text{SORT}(K_y^-, d)$  ▷ order by increasing distance
             $K_y^- \leftarrow \text{TRIM}(K_y^-, N_{max})$  ▷ keep only the closest  $N_{max}$  events
           $v_{kij} \leftarrow \text{null}$  ▷ all voxels are created empty
          if  $K_y^- \neq \emptyset$  then
             $V, W \leftarrow 0$  ▷ loop dummy variables
            for  $x_n \in K_y^-$  do
               $V \leftarrow V + w(x_n, y) v(x_n, y)$ 
               $W \leftarrow W + w(x_n, y)$ 
             $v_{kij} \leftarrow V/W$  ▷ a valid voxel
         $M \leftarrow M \cup \{(y_{kij}, v_{kij})\}$ 
  return  $M$  ▷ the finished model

```

The most involved part of the evaluation algorithm is the selection of the causally connected sample points to be used in the summation loop. The next algorithm is very general and includes, as a matter of principle, the evaluation of the length of the geodesic line Γ connecting x_n and y_{kij} . Though formally correct, this is not what it is actually performed during the calculations, which are specialized according to the spatial geometry.

A null value of d is associated with those events that are outside the backwards causal cone. Null- d elements are sorted down to the tail of K_y^- . The actual calculation of d_s depends on the geometry of the space.

As of now, the Local and Global implementations of the Timescape Algorithm do the following:

- **TimescapeGlobal** employs a spherical approximation of the Earth surface, so that¹⁰²
 $d_s = R \arccos(\sin \varphi_{x_n} \sin \varphi_y + \cos \varphi_{x_n} \cos \varphi_y \cos(\lambda_{x_n} - \lambda_y))$ and the cell size w_c is specified in angular units.
- **TimescapeLocal** uses projected Euclidean coordinates (q, p) to represent the spatial component of the events, so $d_s = \sqrt{(q_{x_n} - q_y)^2 + (p_{x_n} - p_y)^2}$. Alternatively, the equivalent diamond and square metrics can be used (10).

As said, any voxel is evaluated independently from the others, so any model can be easily subdivided according to a union of submodels, provided some care is given to the limits of the submodels and that cell- and sheet sizes are kept unchanged.

A.5 Jackknifing and Ensemble Means

Depending on the consistency of the samples set S , it could be possible to subset jackknife the samples seeking for a bias-corrected estimate of the model values. This is achieved considering S as an *ensemble* of subsets S_k , so that $S = \cup S_k$ and, in general, $S_k \cap S_{k'} \neq \emptyset$. Depending on the consistency of S this procedure can be performed or not (all S_k must be statistically significant). The standard Jackknifing procedure consists in correcting a biased estimate¹⁰³ using a collection of subsets S_k , each of which neglects only the element x_n :

Algorithm 3 Model Jackknifing

```

procedure JACKKNIFE( $S$ )
   $M = \text{MODEL}(S)$  ▷ Evaluate the global model
  for  $n \leftarrow 1 \dots N$  do
     $M_n \leftarrow \text{MODEL}(S \setminus \{x_n\})$  ▷ Evaluate the  $n$ th model
  for  $y_{kij} \in M$  do
     $\hat{\theta}_{kij} \leftarrow v_{kij}$  ▷ Global estimator
     $\hat{\theta}_{kij}^* \leftarrow 0$  ▷ Biased estimator
    for  $n \leftarrow 1 \dots N$  do
       $\hat{\theta}_{kij}^* \leftarrow \hat{\theta}_{kij}^* + v_{kij}^{(n)} / N$ 
     $\hat{\theta}_{kij} \leftarrow N \hat{\theta}_{kij}^* + (1 - N) \hat{\theta}_{kij}^*$  ▷ Correct the estimator of the global model
  return  $M$ 

```

Jackknifing adds another factor N to the complexity of the calculations.

Another smart trick allowed by ensemble techniques allows a sort of “reverse modeling”. If one knows not how to assign a value to the c and k parameters it is possible to use the Timescape

¹⁰²Which is the spherical version of $\int_{\Gamma} \sqrt{g} d\tau$ evaluated on a path Γ between $\mathbf{x}_n$ and $\mathbf{y}$, the spatial components of the events $x_n = (t_{x_n}, \mathbf{x}_n)$ and $y = (t_y, \mathbf{y})$.

¹⁰³This does not correct a biased set of observations, of course.

Algorithm in order to find an estimate of such parameters. Since c and k are related to the transport/diffusion capabilities of the system, i.e. to the patterns of change of the investigated phenomenon, what we gain is, in fact, an estimate of such velocity.

We can think $(c, k) \in \mathbb{R}^+ \times \mathbb{R}^+$ as a space of parameters, for each (c, k) pair¹⁰⁴ we have a distinct behaviour of the model, it is possible that some values of c and k match the actual patterns of change better than others. To find these values it is customary to select a *control group* subsetting S into two disjoint sets: a sample set S_0 and a control set S_c .

Then we interpolate a set of models $\{M_n\}$, each of which corresponds to a parameters pair (c_n, k_n) . From the model M_n we then evaluate the residuals according to S_c and try to minimise them. This is done evaluating the squares of the differences between the elements of S_c and the model-estimated values at the same events.

One thing we can try is a recursive adjustment of the parameters down to an error ξ :

Algorithm 4 Recursive Parameters Estimate

```

procedure ESTIMATE( $S_0, S_c, \xi$ )
   $X_c \leftarrow \{(x_n^c, v_n^c) \in S_0\}$  ▷ The set of control events and associated values
   $N \leftarrow \#(S_c)$ 
   $E \leftarrow +\infty$  ▷ A trivially high error value
  while  $E > \xi$  do
    ADJUST( $c, k$ ) ▷ Find new parameters
     $M \leftarrow \text{MODEL}(S_0)$ 
     $E \leftarrow 0$ 
    for  $x_n^c \in X_c$  do
       $v_n \leftarrow v_{kij} \mid d(y_{kij}, x_n^c) = \min$  ▷ Estimated value
       $E \leftarrow E + (v_n^c - v_n)^2$ 
     $E \leftarrow \sqrt{E}/N$ 
  return  $M$ 

```

This is an acceptable procedure if there is any clue about an ADJUST(c, k), procedure otherwise it is just a random wandering in the parameters space. This is where *ensemble averages* come into play. We create a finite set P of tentative pairs of parameters $(c_n, k_n) =: p_n$ each of which corresponds to a model M_n (given the S_0 and S_c sets). We then evaluate all the models and the associated errors E_n .

Supposing that P exhaustively represents all the possible cases.¹⁰⁵ This can be extended to different pairs of positive functions $(c(x, y, v, \mathbf{a}), k(x, y, v, \mathbf{a}))$ which is a natural extension of simple constant values, but it complicates the already complex calculations beyond the reach of ordinary desktop computers.

¹⁰⁴Let's limit ourselves to the case of constant parameters. The general idea can be applied also to functions, but the complexity grows accordingly.

¹⁰⁵Here is *Ergodicity* hidden beneath the *exhaustively represents*. See e.g. [Sethna 2006].

The estimate goes as follows:¹⁰⁶

Algorithm 5 Ensemble Parameters Estimate

```

procedure ESTIMATE( $S_0, S_c, P$ )
   $\tilde{c}, \tilde{k}, w \leftarrow 0$                                  $\triangleright$  Define new parameters and a global weight
   $X_c \leftarrow \{(x_n^c, v_n^c) \in S_c\}$                  $\triangleright$  The set of control events and associated values
   $N \leftarrow \#(S_c)$ 
  for  $(c_n, k_n) \in P$  do
     $M_n \leftarrow \text{MODEL}(S_0, c_n, k_n)$ 
     $e_n \leftarrow 0$ 
    for  $x_k^c \in X_c$  do
       $v_k \leftarrow v_{kij} \mid d(y_{kij}, x_k^c) = \min$        $\triangleright$  Estimated value
       $e_n \leftarrow e_n + (v_k^0 - v_k)^2$ 
       $w_n \leftarrow N / \sqrt{e_n}$                          $\triangleright$  A weight
       $w \leftarrow w + w_n$ 
       $\tilde{c} \leftarrow \tilde{c} + w_n c_n$ 
       $\tilde{k} \leftarrow \tilde{k} + w_n k_n$ 
     $\tilde{c} \leftarrow \tilde{c} / w$ 
     $\tilde{k} \leftarrow \tilde{k} / w$ 
     $\tilde{M} \leftarrow \text{MODEL}(S_0, \tilde{c}, \tilde{k})$                  $\triangleright$  Ensemble Model
  return  $\tilde{M}$ 

```

Also this procedure adds a significant amount of complexity to the calculations ($N + 1$ models need to be calculated).¹⁰⁷ Nonetheless, the added benefit of obtaining an estimate of the dynamic parameters could shed a light on an otherwise obscure phenomenon. It is possible, however, to evaluate N downscaled, tiny models and then the full-scaled $\tilde{M}$ with the same parameters values.

It is advisable not to try to build a “smart set” of tentative parameters pairs. If one has an idea about their value, it is better to start a recursive seek.

Those interested in the genesis of the algorithm can find the details of spacetime structure in [Frankel 2012]. General spacetime variables techniques are described in [Szeckeres 2006]. It is particularly important to see how these Quantum Field techniques should be transferred in the classical domain [Schlosshauer 2008].

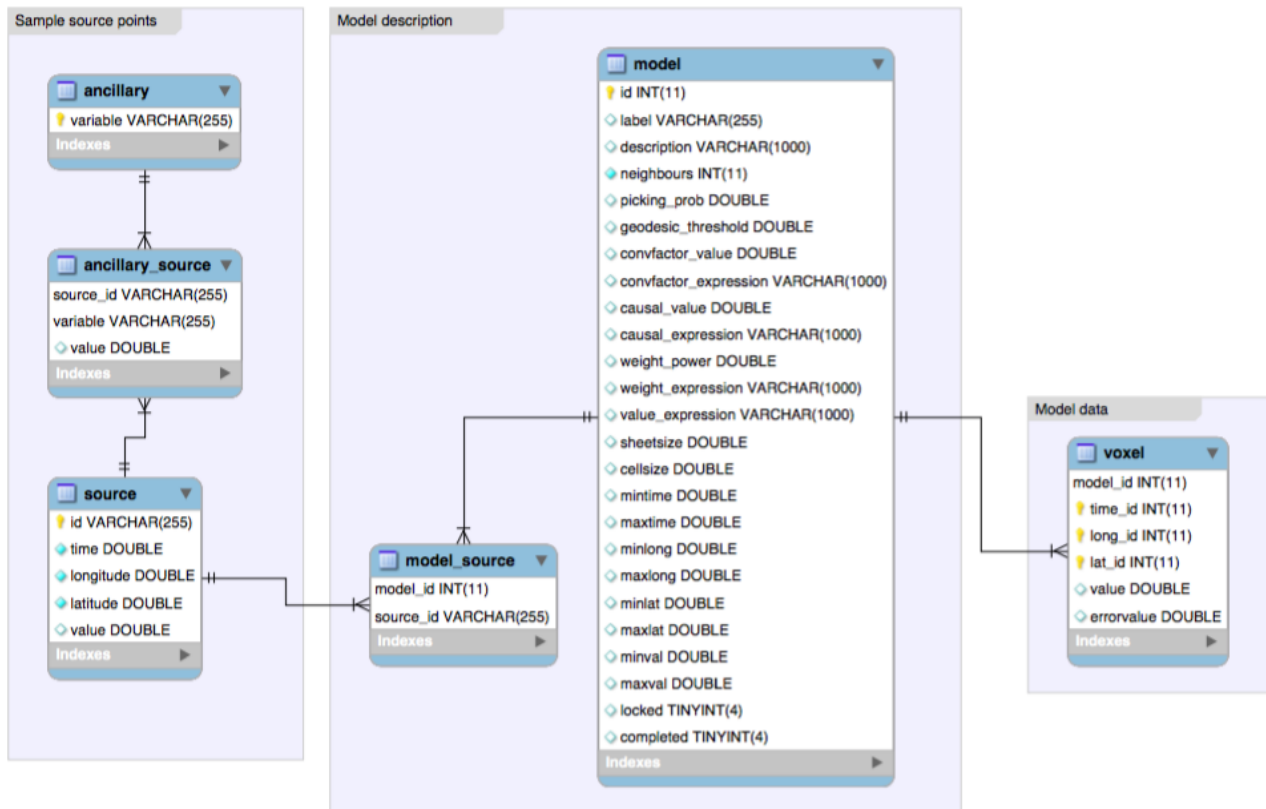
¹⁰⁶In this case, S_0 and S_c need not be disjoint; as a limiting case they can be both S .

¹⁰⁷As is always the case with statistical ensembles, N should be a really big number.

B Appendix: The Timescape Data Structure

This appendix describes the **TimescapeGlobal** data structure, which is neater and easier to understand than the local version. It is documented in detail in the documentation accompanying the software distribution.

The database structure is pretty simple. Three main blocks can be found; a Source block, a Model description and a Model data block:



- Source: Three tables contain the samples data values. The main table is **source**, which stores the spacetime coordinates and the values of the sample points.¹⁰⁸ Two other tables, **ancillary** and **ancillary_source** are used only when ancillary data are collected.
- Model: a single **model** table stores all the relevant parameters to each model. Fields include all the descriptors of the model, plus the minimum and maximum values of the evaluated voxels and a couple of flags used to record whether the model is being evaluated or already completed. The **model_source** table is a cross-reference object that stores which source point are actually used in each model.¹⁰⁹

¹⁰⁸Values cannot be null.

¹⁰⁹There is not a global on/off switch on source points.

- Data: a single **voxel** table is used to store the model voxels. To keep the record size to a minimum, the actual coordinates are not stored in the table. Only the coordinate indices are stored. The *errorvalue* field is as yet unused.

Although the application is devoted to geographical data, it is not important to have spatially indexed fields. Care is given to keep the number database queries to a minimum and a gain in efficiency is advisable only for the exploration of completed models. During the evaluation of the models, empty records are inserted prior to the actual calculation to avoid the continuous grow of **voxel** table and to be sure that the evaluation will be not aborted due to lack of database space.

The user can adopt the database flavour and location to taste.¹¹⁰ Connection speed is of the utmost importance, mostly during the phases of voxels insertion and of model exploration (the latter, consisting in select-only accesses, is far less critical then the insertion phase).

The detailed structure of the tables follows:

Table ancillary				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	variable	String	varchar(255)	the ancillary variable name

Table source				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	id	String	varchar(255)	the sample name
	time	double	double	sample time, any unit
	longitude	double	double	sample longitude, d.ddd
	latitude	double	double	sample latitude, d.ddd
	value	double	double	the sample value (cannot be null)

Table ancillary_source				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	source_id	String	varchar(255)	sample id → source
◇	variable	String	varchar(255)	variable → ancillary
	value	double	double	the ancillary variable value (can be null)

¹¹⁰The ER model comes from a MySQL implementation of the data model.

Table model				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	id	int	int(11)	self-incremented model id
	label	String	varchar(255)	user-defined model label
	description	String	varchar(1000)	user-defined model description
	neighbours	int	int(11)	number of neighboUring points
	picking_prob	double	double	point picking probability
	geodesic_threshold	double	double	maximum allowed geodesic distance
	convfactor_value	double	double	time to space conversion factor
	convfactor_expression	String	varchar(1000)	time to space conversion expression
	causal_value	double	double	causal cone ratio
	causal_expression	String	varchar(1000)	causal cone ratio expression
	weight_power	double	double	interpolation weight function power
	weight_expression	String	varchar(1000)	interpolation weight function expression
	value_expression	String	varchar(1000)	interpolation value function expression
	sheetsize	double	double	time sheet interval, any unit
	cellsize	double	double	space cell size, d.ddd
	mintime	double	double	minimum time, any unit
	maxtime	double	double	maximum time, any unit
	minlong	double	double	minimum longitude, d.ddd
	maxlong	double	double	maximum longitude, d.ddd
	minlat	double	double	minimum latitude, d.ddd
	maxlat	double	double	maximum latitude, d.ddd
	minval	double	double	minimum value, any unit
	maxval	double	double	maximum value, any unit
	locked	boolean	tinyint(4)	model locked flag
	completed	boolean	tinyint(4)	model complete flag

Table model_source				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	model_id	int	int(11)	model id → model
◇	source_id	String	varchar(255)	sample id → source

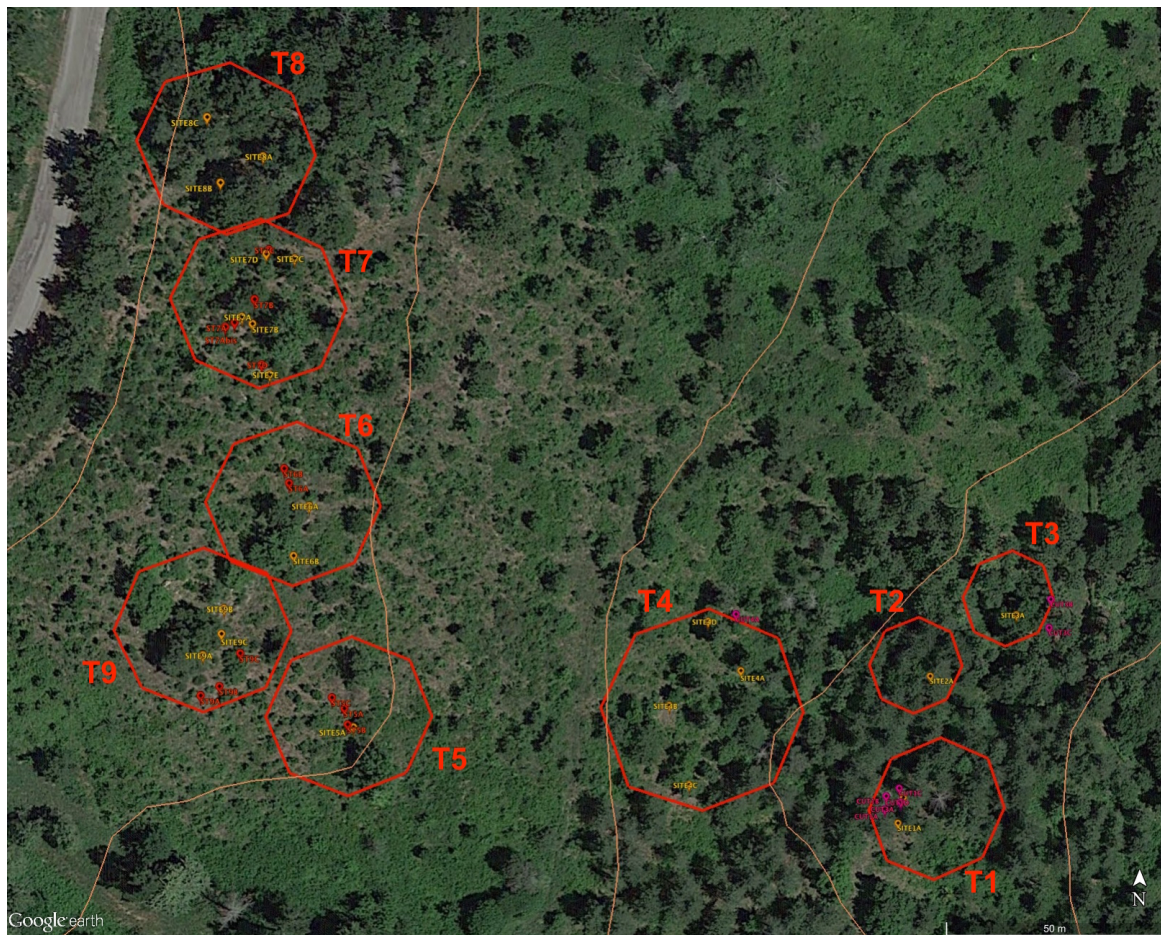
Table voxel				
INDEX	FIELD	JAVA TYPE	MYSQL TYPE	DESCRIPTION
◇	model_id	int	int(11)	model id → model
◇	time_id	int	int(11)	time sheet id
◇	long_id	int	int(11)	longitude cell id
◇	lat_id	int	int(11)	latitude cell id
	value_id	double	double	voxel value, can be null
	errorvalue_id	double	double	voxel value, can be null, unused

It is worth noting the logical structure of the tables. **model** stores all the informations relevant to the model, but not the model itself, which is stored in the **voxel** table. The latter has not room for the actual coordinates, only integers pointers to them are recorded, to keep the table size to a minimum. **voxel** stores the bulk of the model and it is the only table that can inflate, giving storage problems during the usage of the software. On the other hand, keeping the model data on a single table allows for an easy export of the data according to users' needs.

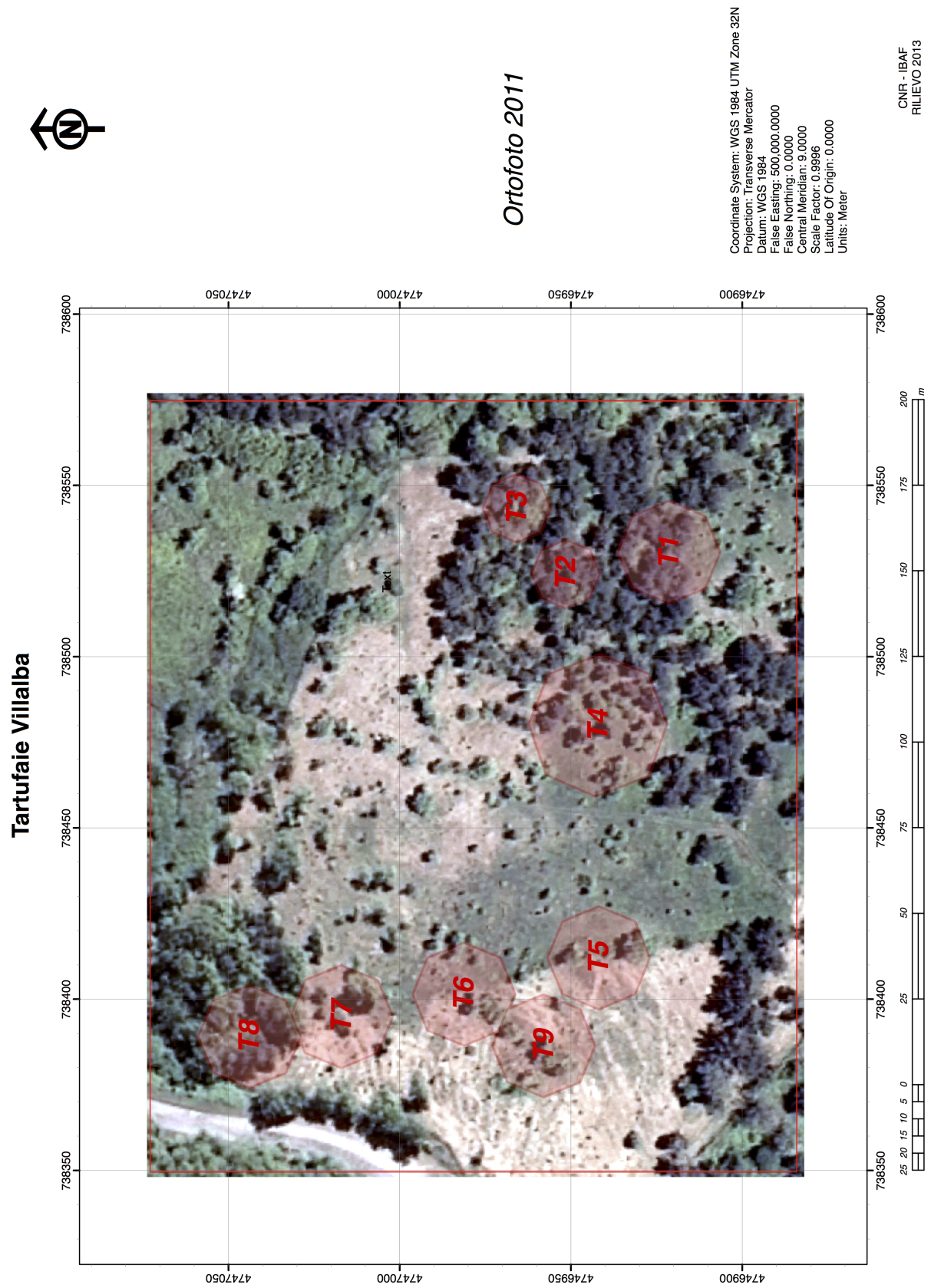
C Appendix: Mycorrhiza Survival Strategy Sampling Atlas

This appendix contains the detailed maps of all the collected samples of the Symbiosis case study reported on section 4.

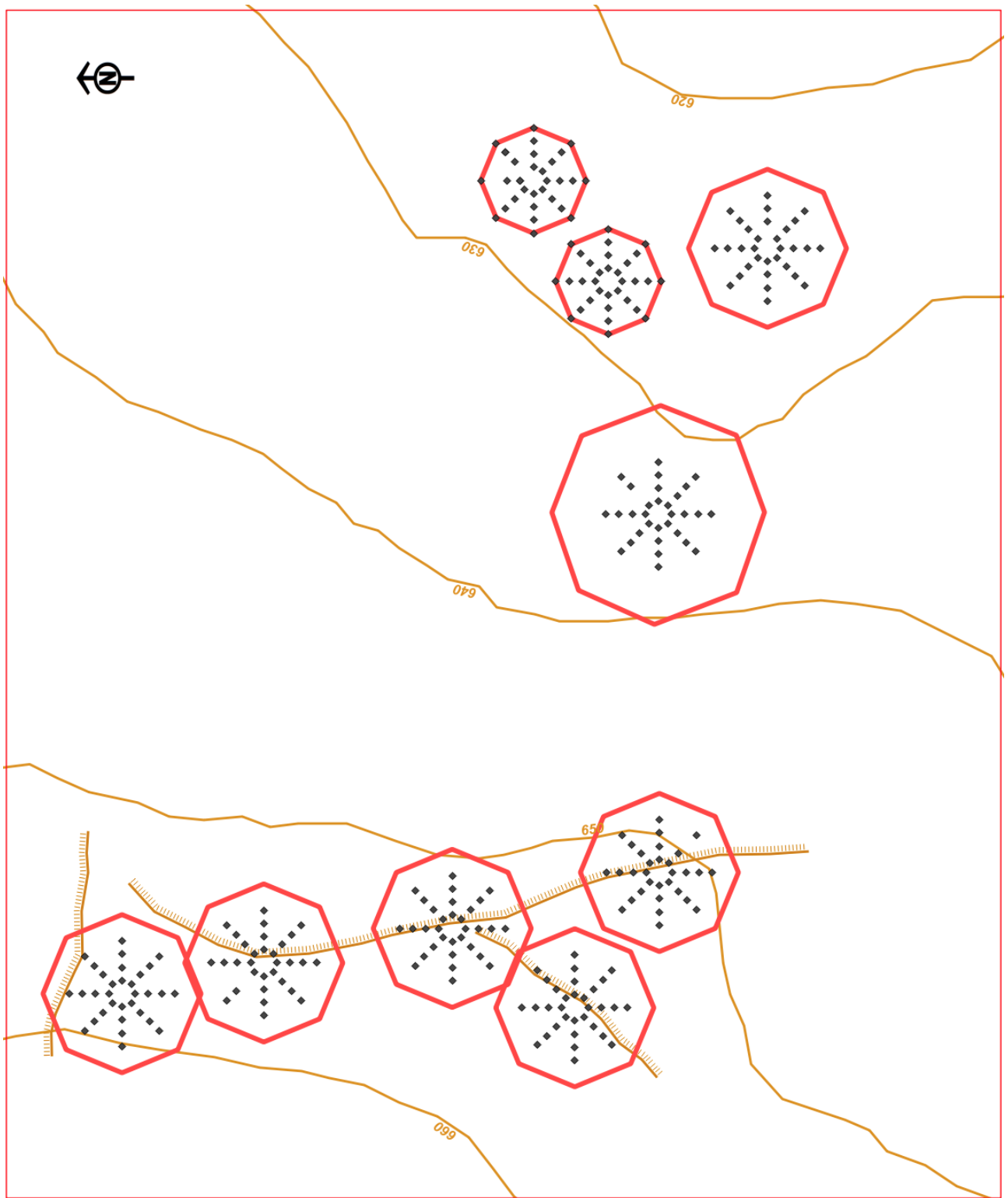
- a general orthophotograph of the whole area, depicting the nine collection sites
- the soil sampling sites
- the black truffles collection sites
- the stumps sites
- the trees positions, all in a map and divided by species
- the most recent cuts locations



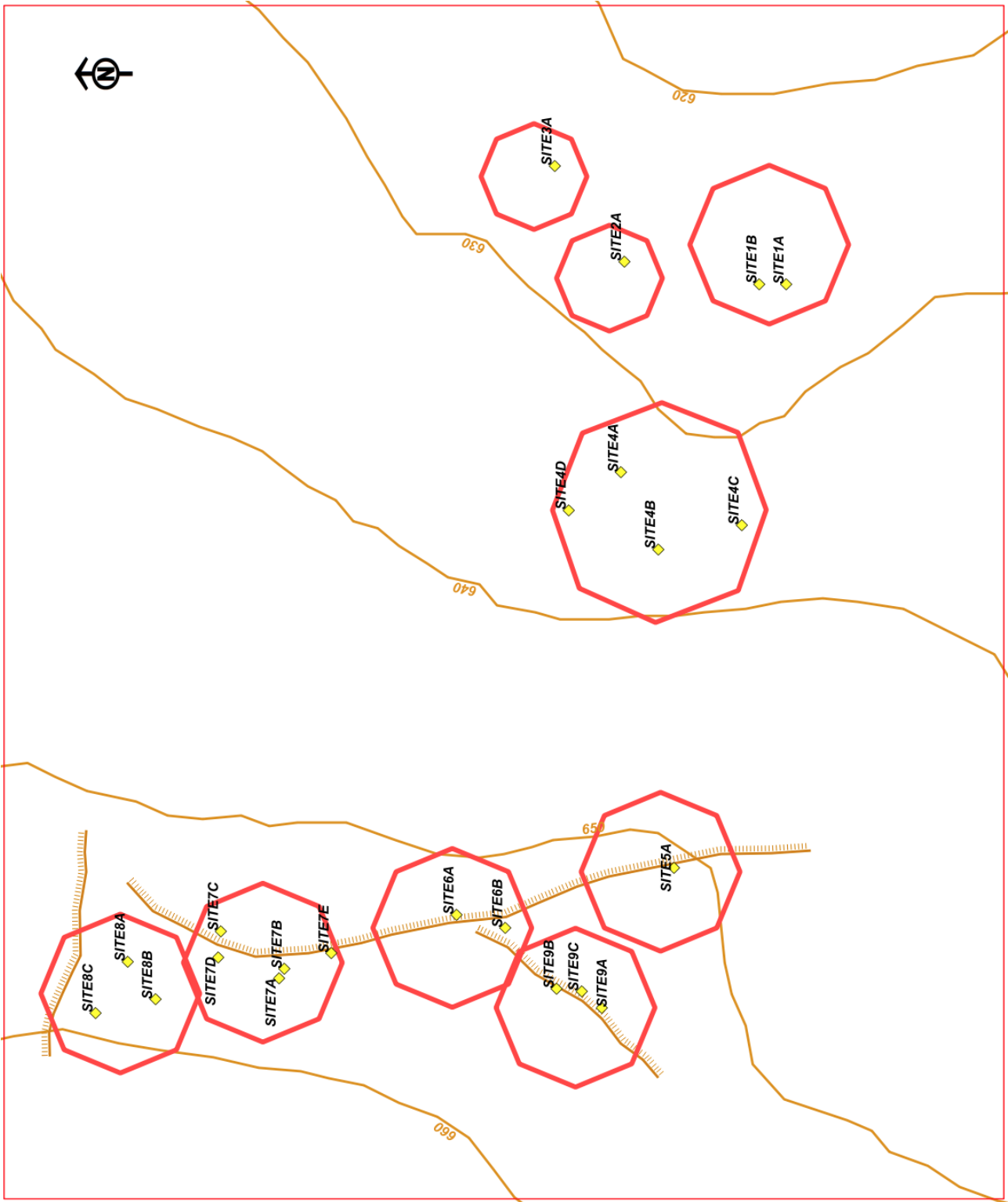
Actual view of the area (2015, from GoogleEarthTM).



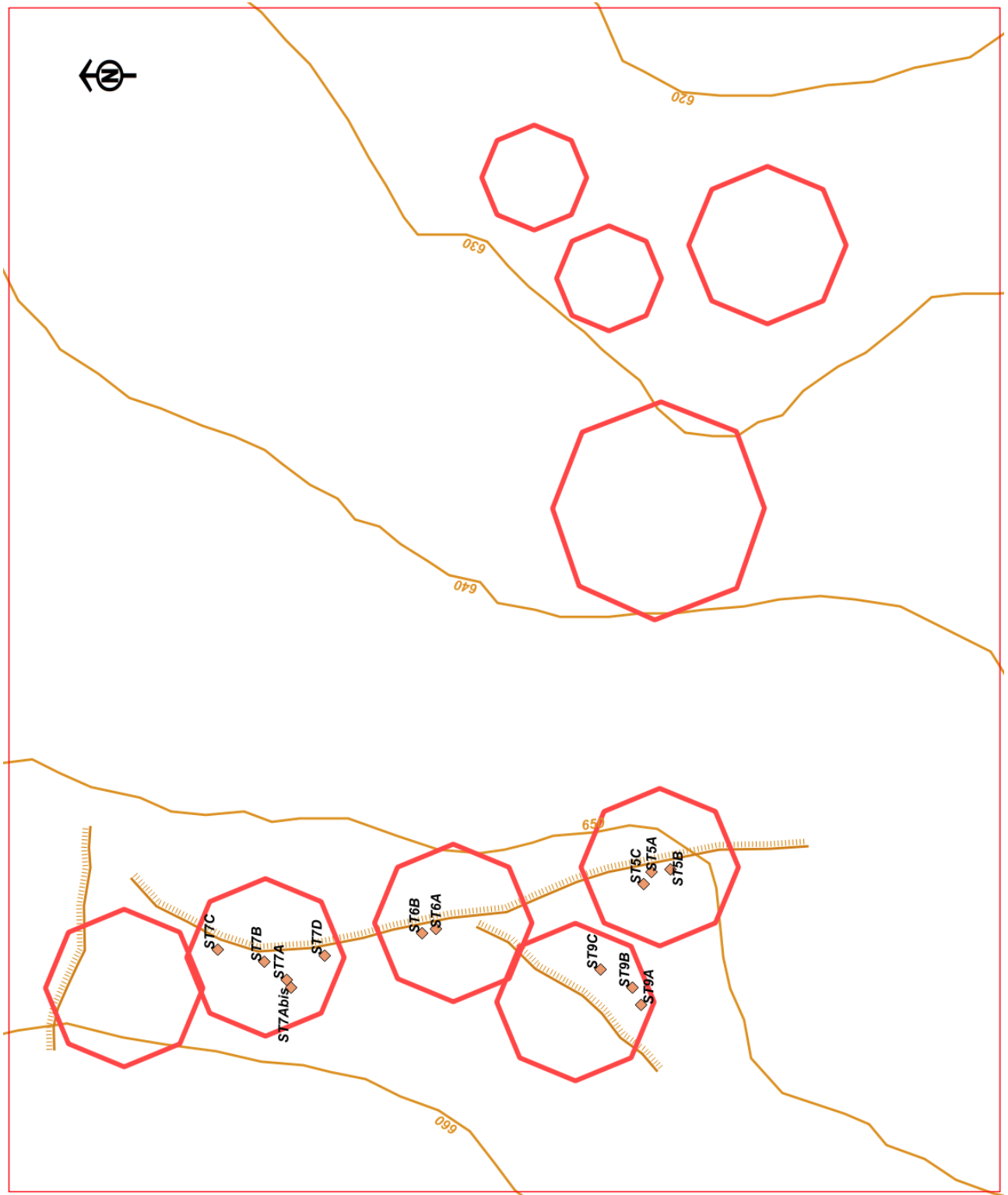
Area Orthophotograph, UTM zone 32 North coordinates referred to WGS84 Ellipsoid



Soil sampling sites



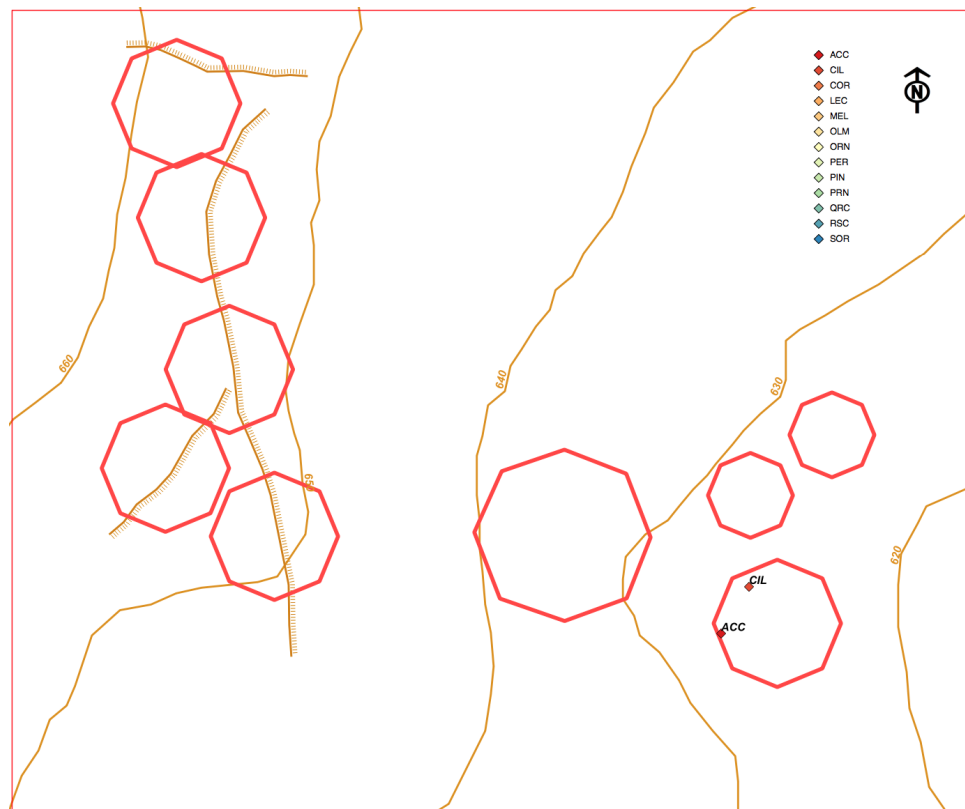
Actual truffles collection spots



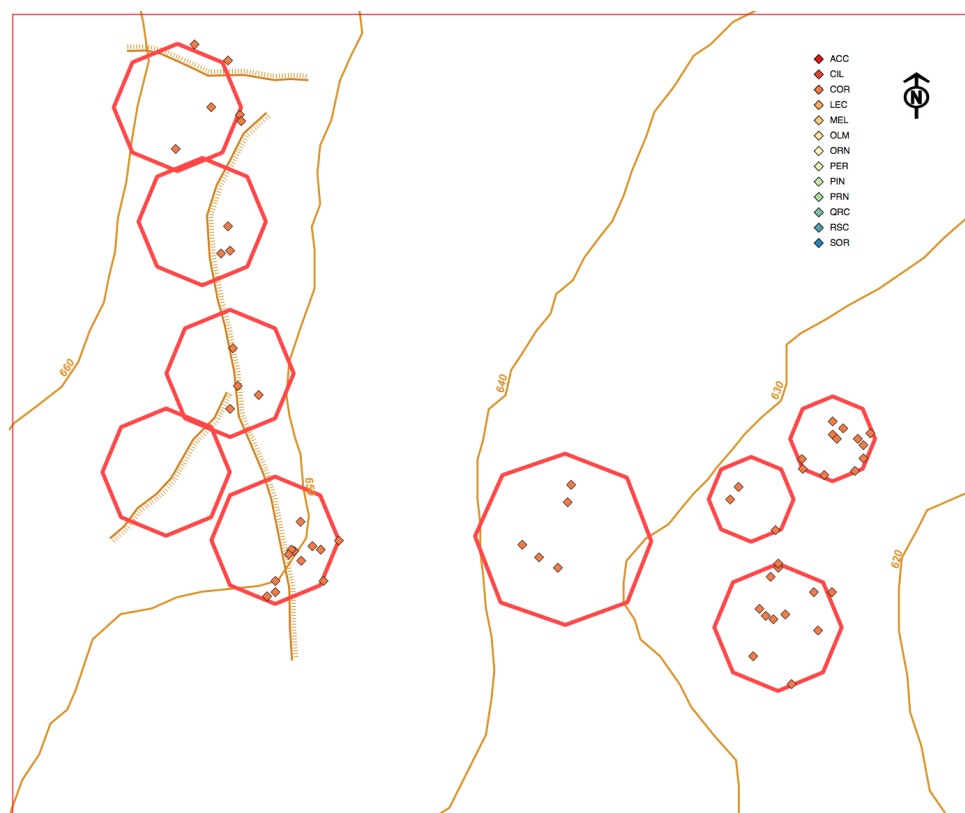
Stumps sites



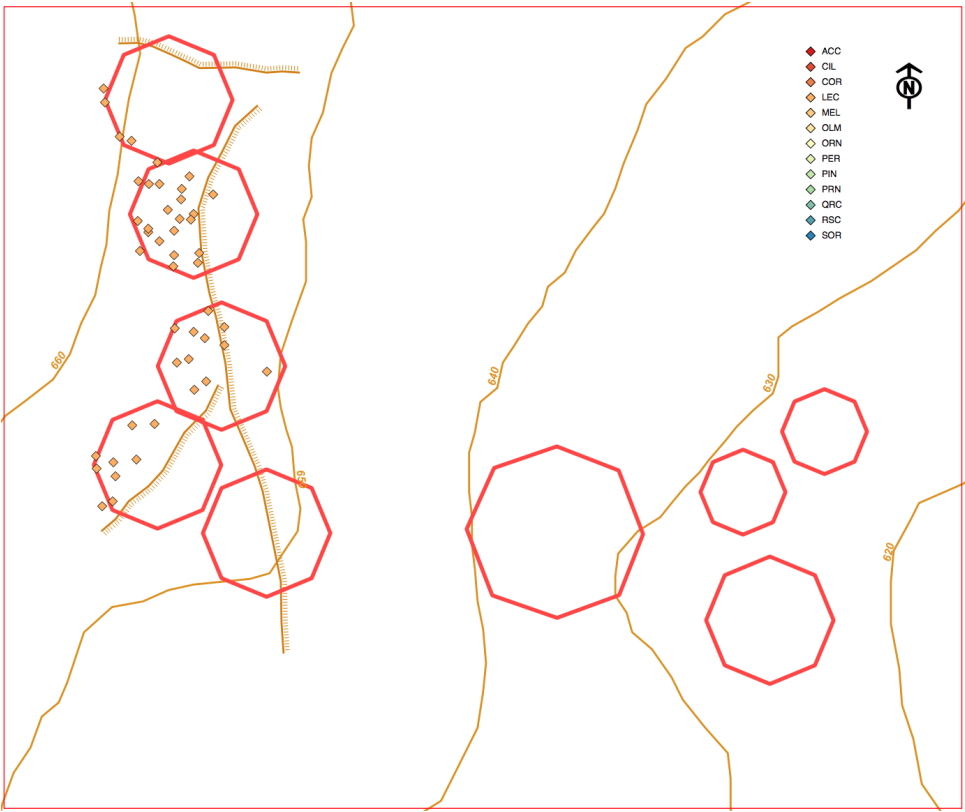
Actual trees locations: ACC *Acer campestre*, CIL *Prunus avium*, COR *Cornus mas*, LEC *Quercus ilex*, MEL *Malus sylvestris*, OLM *Ulmus spp*, ORN *Fraxinus ornus*, PER *Pyrus pyraister*, PIN *Pinus spp*, PRN *Prunus spp*, QRC *Quercus spp*, RSC *Rosa canina*, SOR *Sorbus aucuparia*



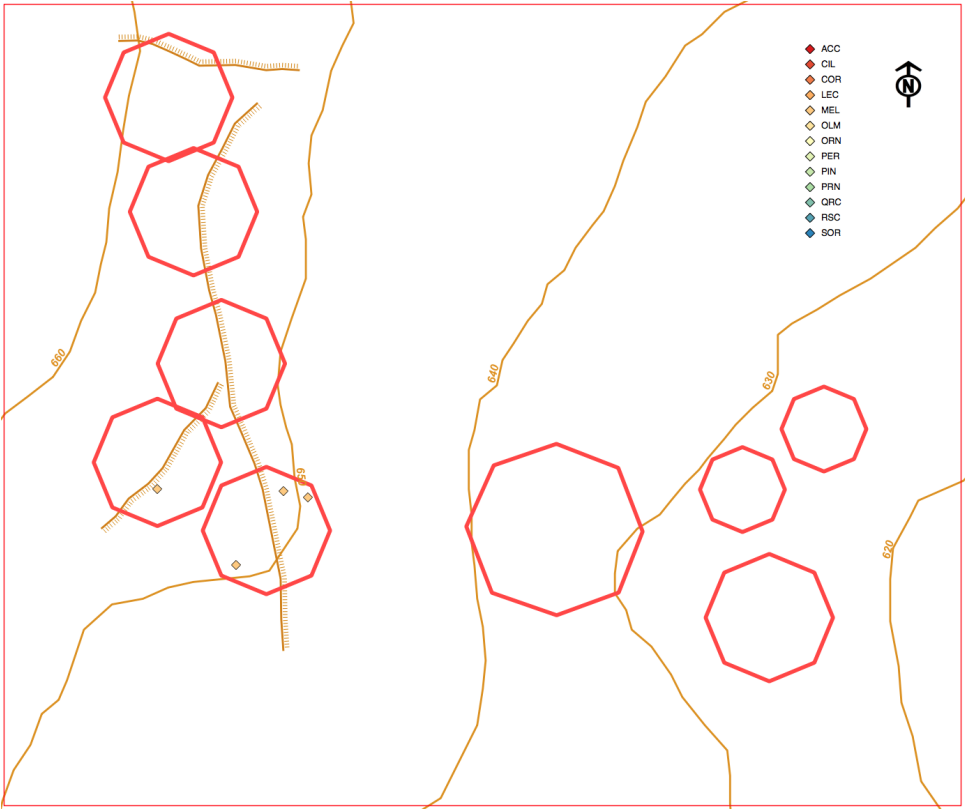
Acer campestre and *Prunus avium* trees location



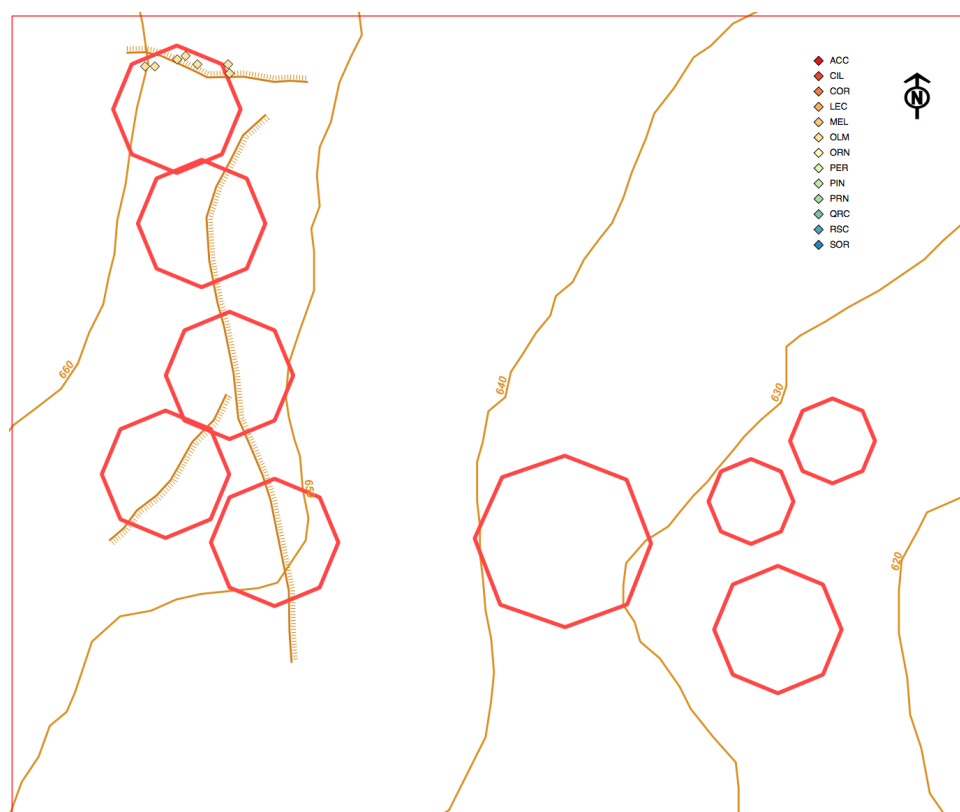
Cornus mas trees location



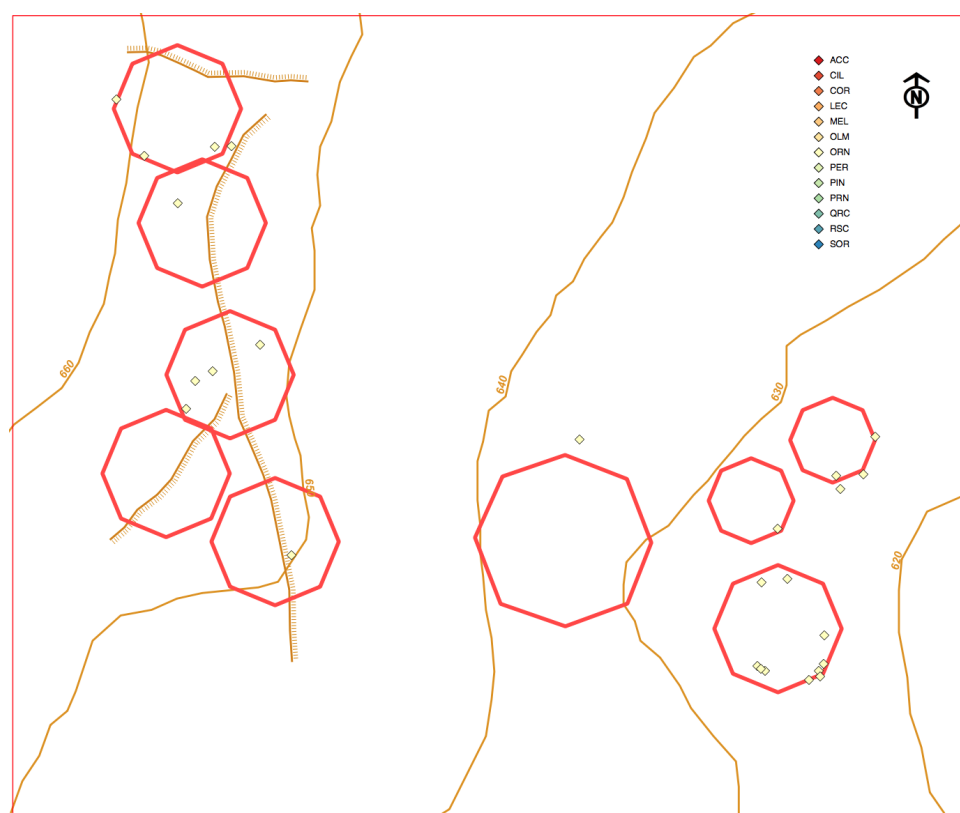
Quercus ilex trees location



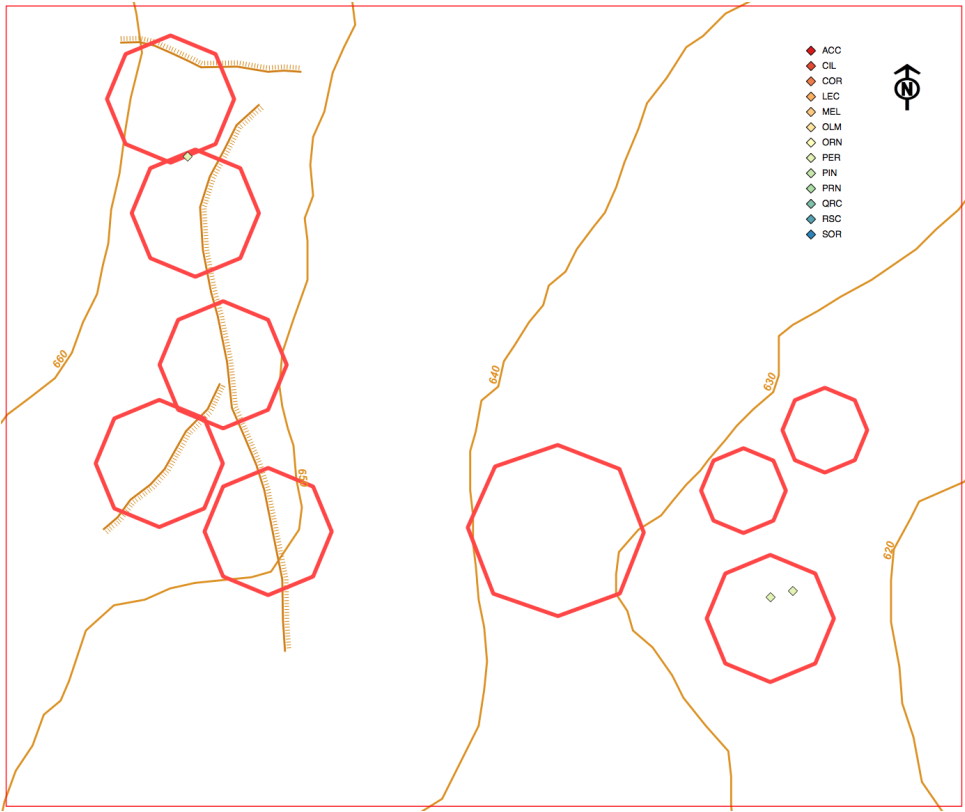
Malus sylvestris trees location



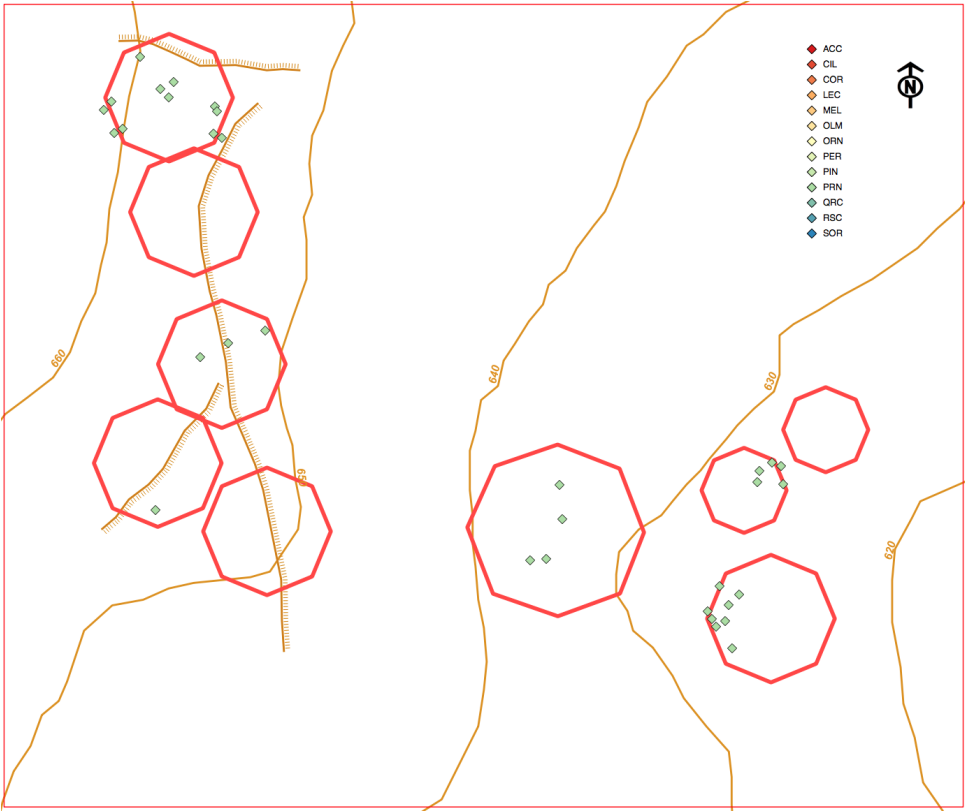
Ulmus spp trees location



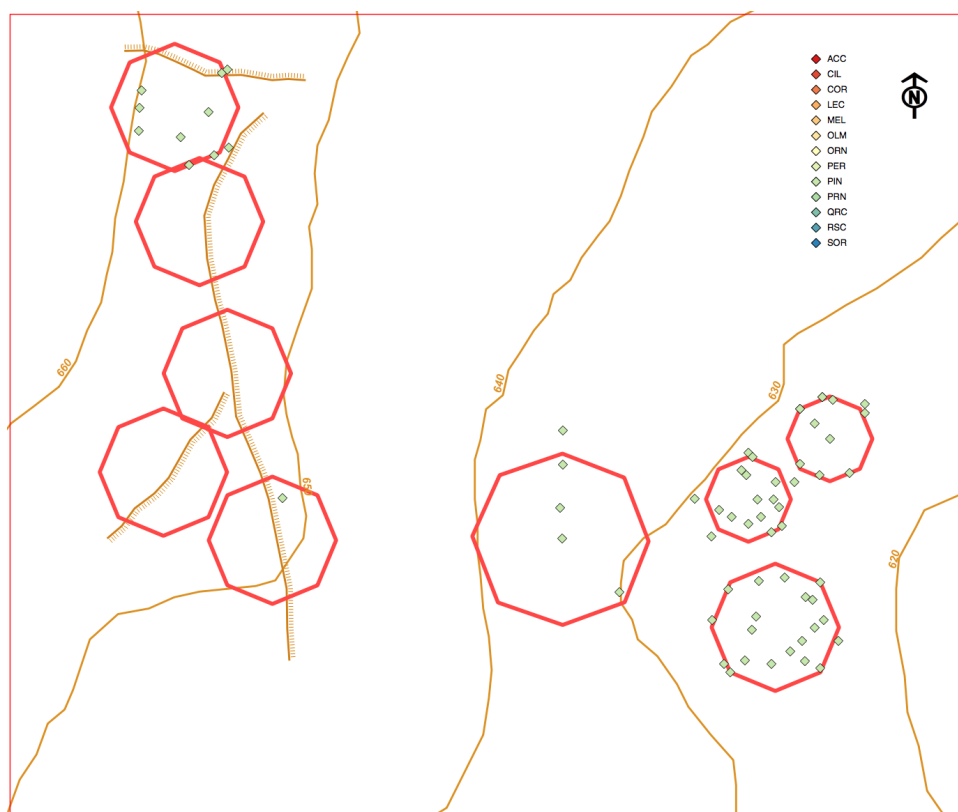
Fraxinus ornus trees location



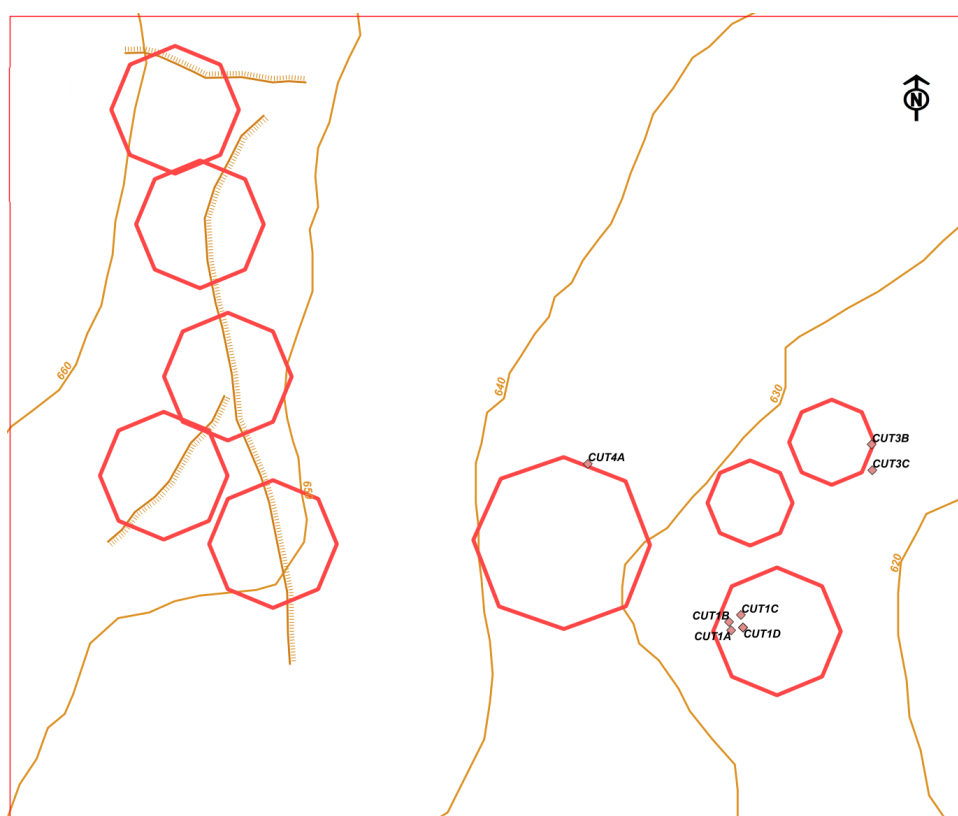
Pyrus pyraeaster trees location



Prunus spp trees location



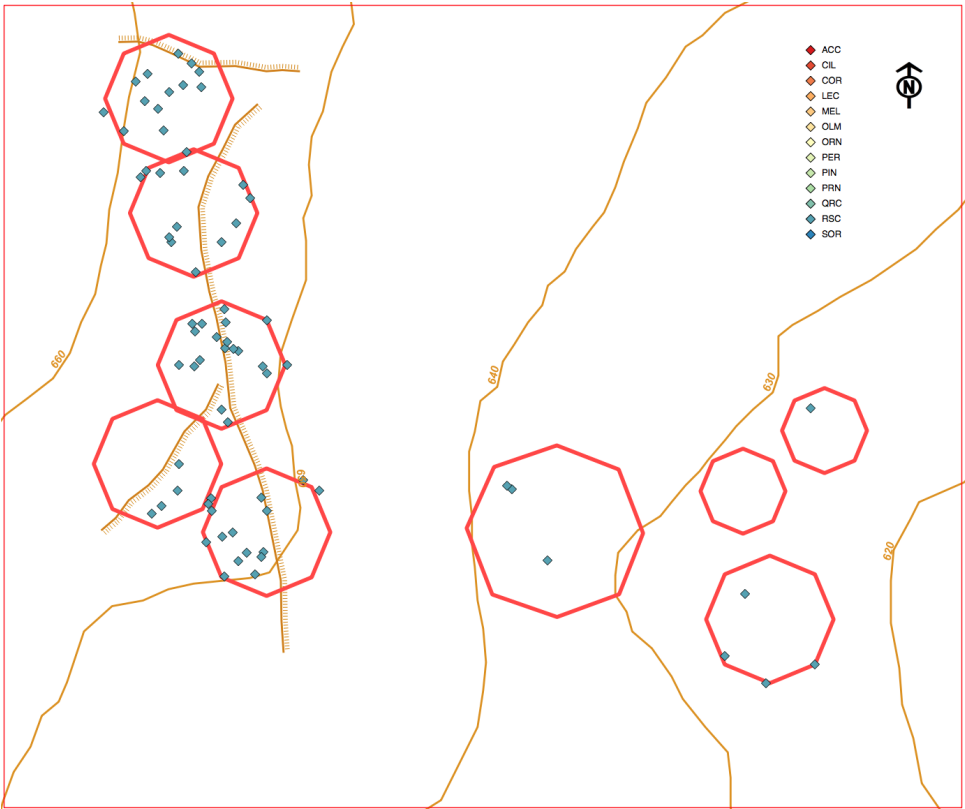
Pinus spp trees location



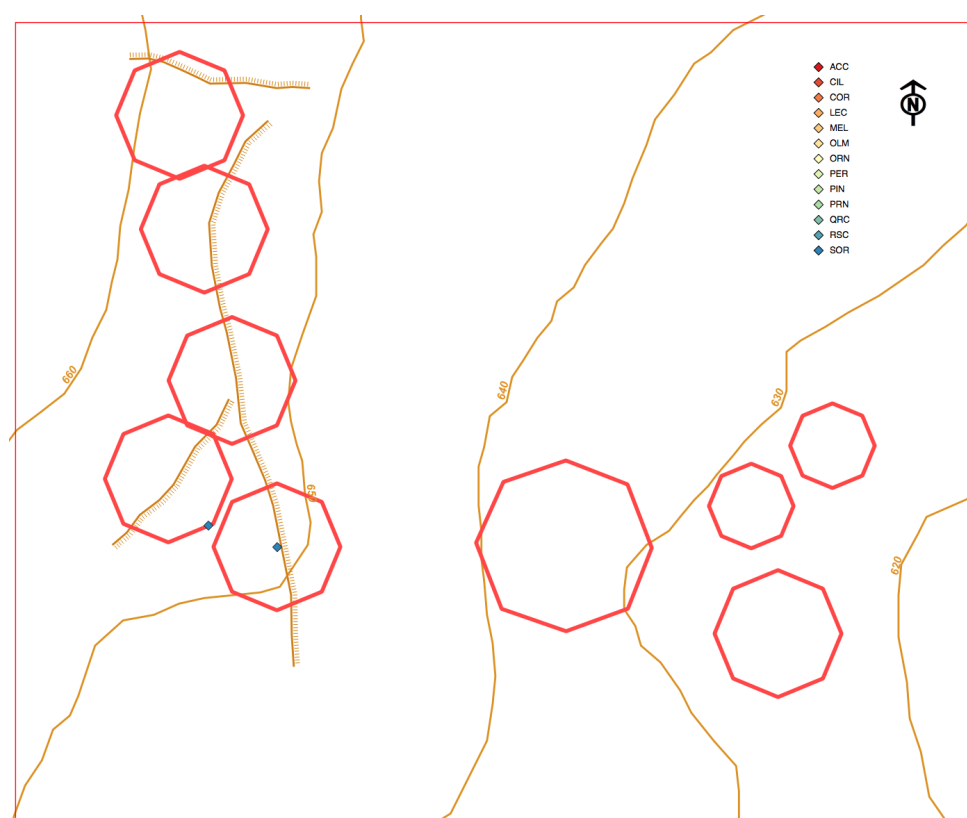
Most recent pine trees cuts (2012)



Quercus spp (other than *Quercus ilex*) trees location

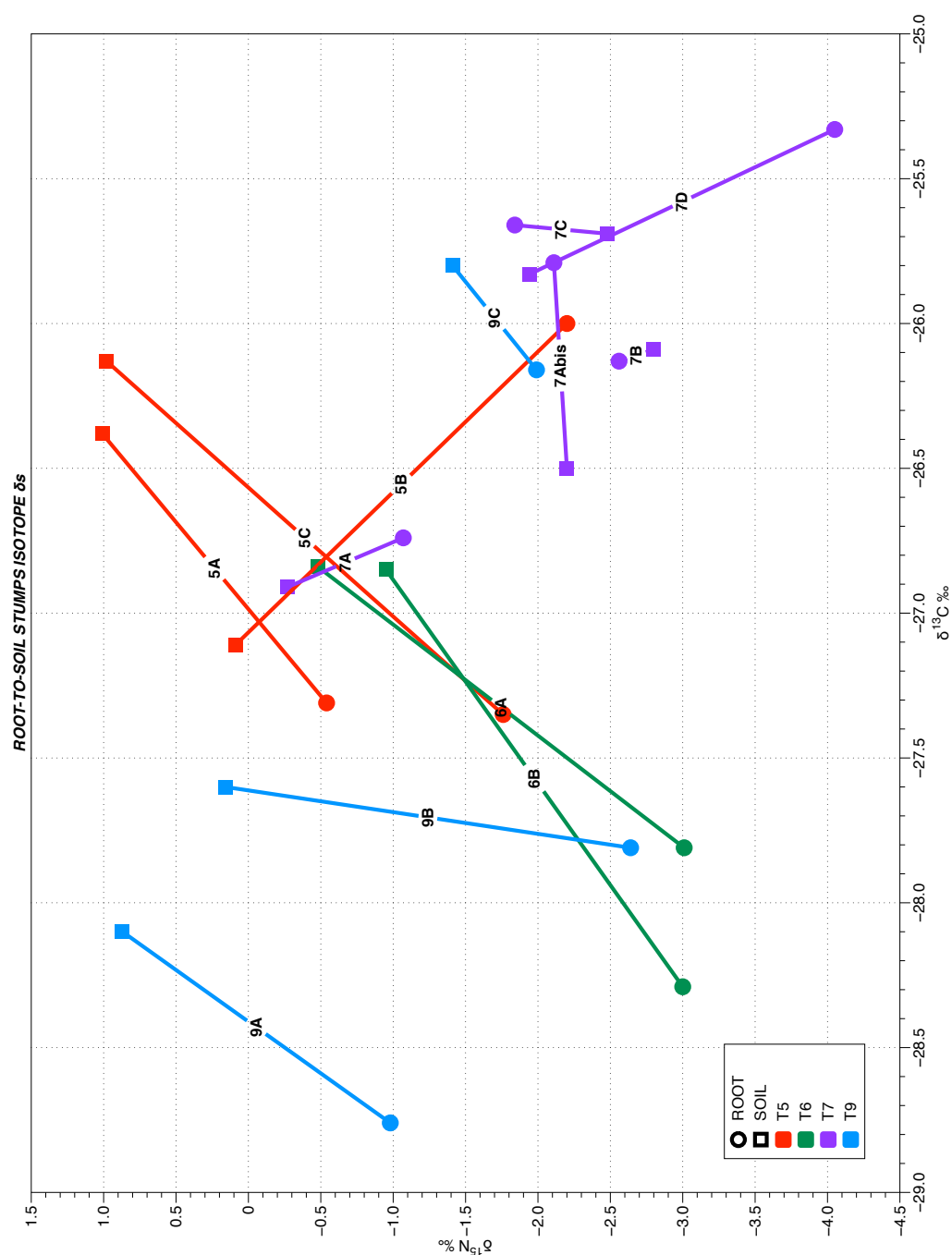


Rosa canina plants location

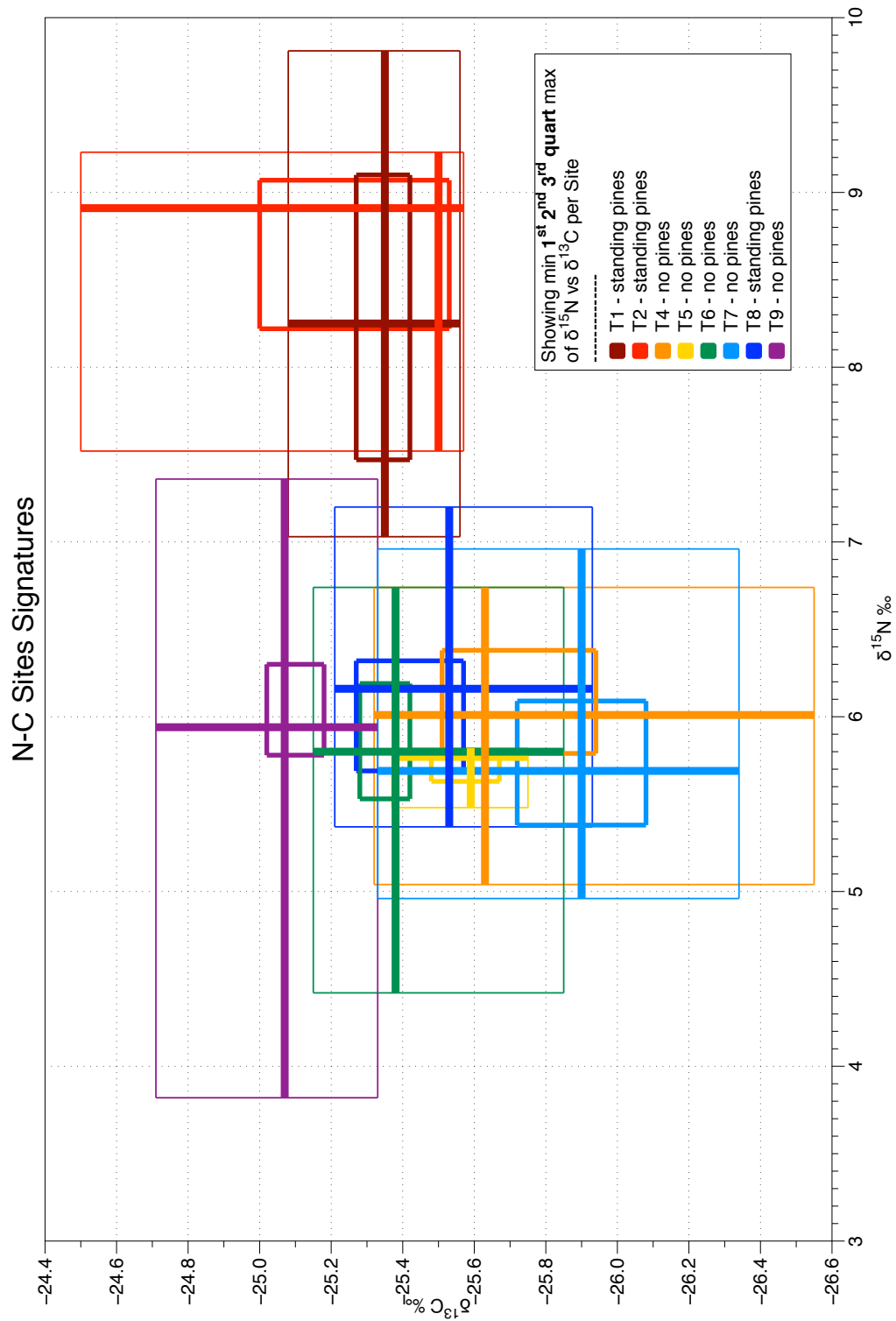


Sorbus aucuparia trees location

D Appendix: Mycorrhiza Survival Strategy Statistics



Values of $\delta^{13}C$ vs $\delta^{15}N$ of the stumps. The circles mark the actual stump tissues, while the squares mark a control analysis performed on the soil nearby. Older stumps (from site T7) have closer soil-stump values.



Distribution of $\delta^{13}\text{C}$ vs $\delta^{15}\text{N}$ of truffles per site. The external box sets the minimum and maximum values, the internal box is proportional to the second quartile (median), while the crosses are located at the mean values. Sites T1 and T2 single out naturally, being more enriched in ^{15}N with respect to the others, this is partially due to the soil bias, that has been accounted for in the modeling. the $\delta^{13}\text{C}$ average values lie within an interval of less than 1‰, while the medians lie within an interval of 1.2‰; this makes the ^{13}C isotope statistically less sensitive.

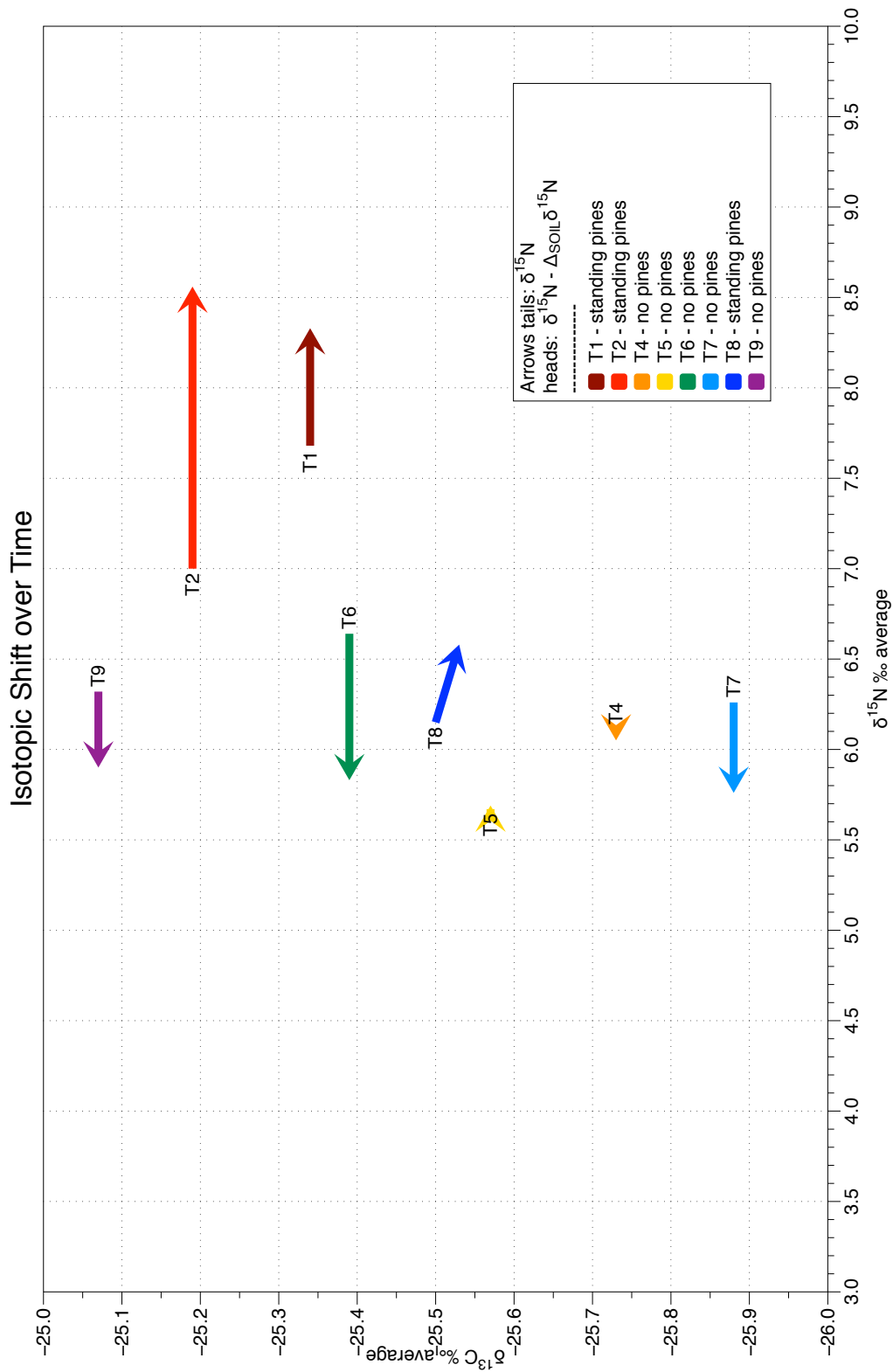
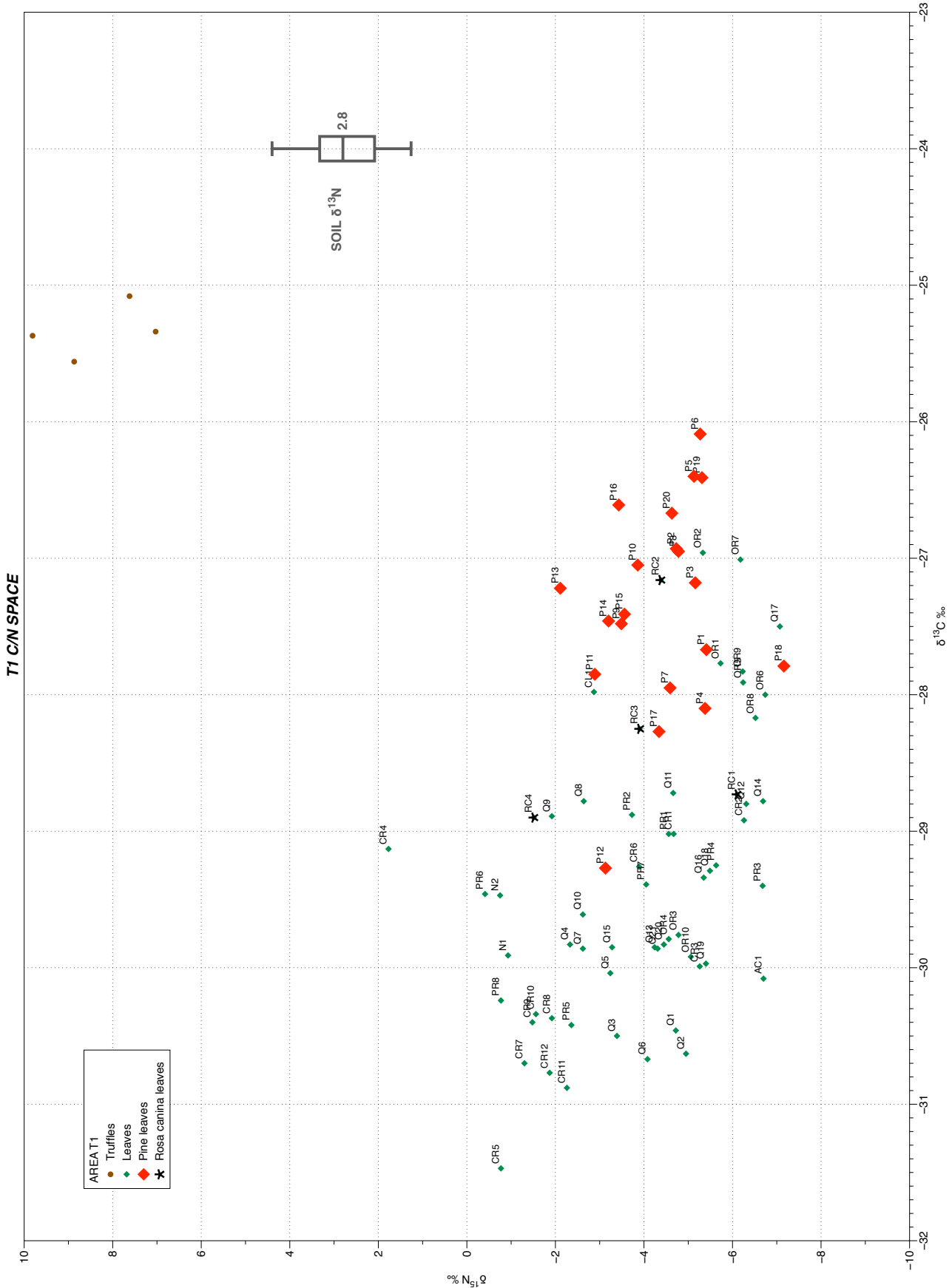
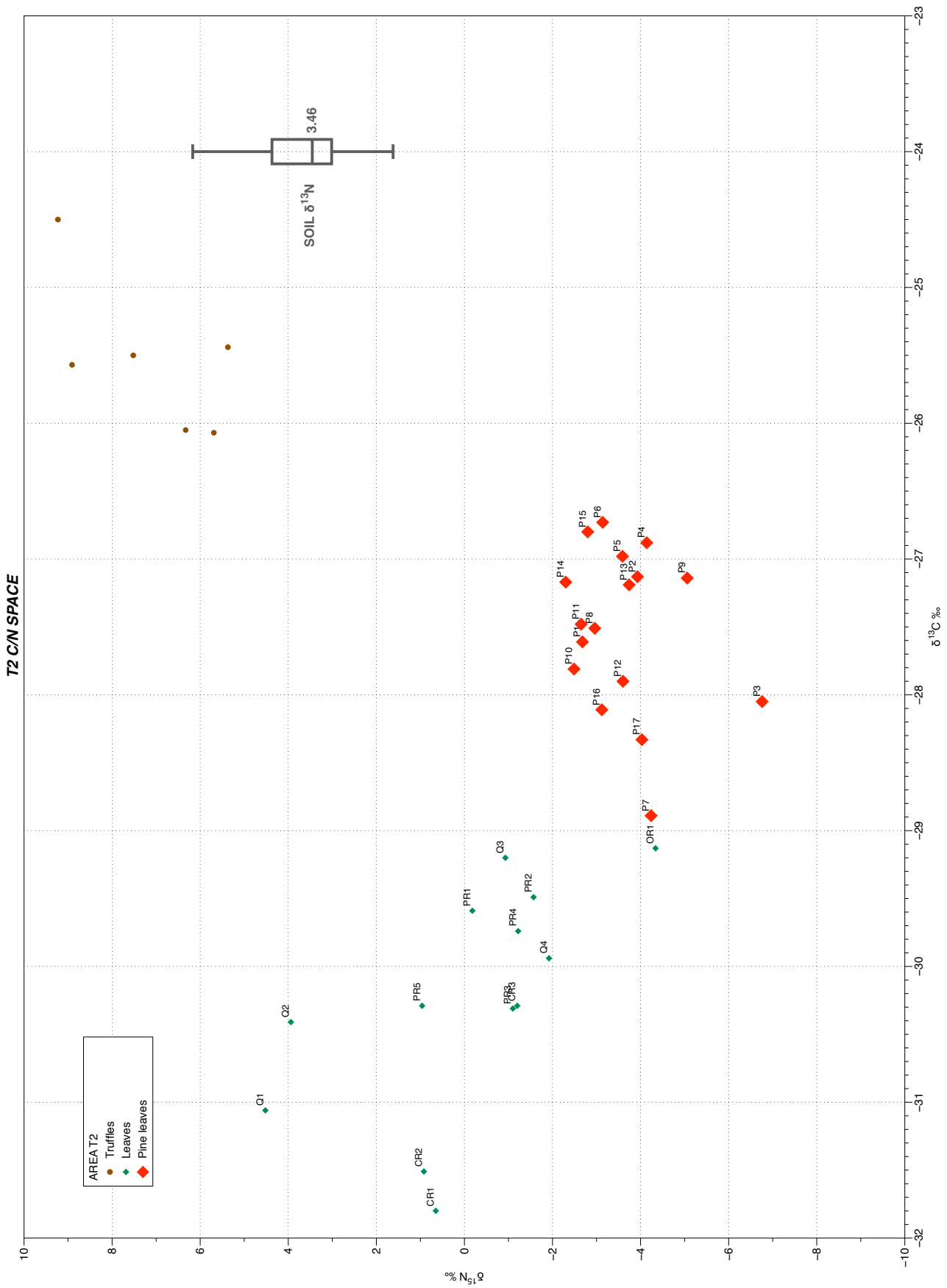


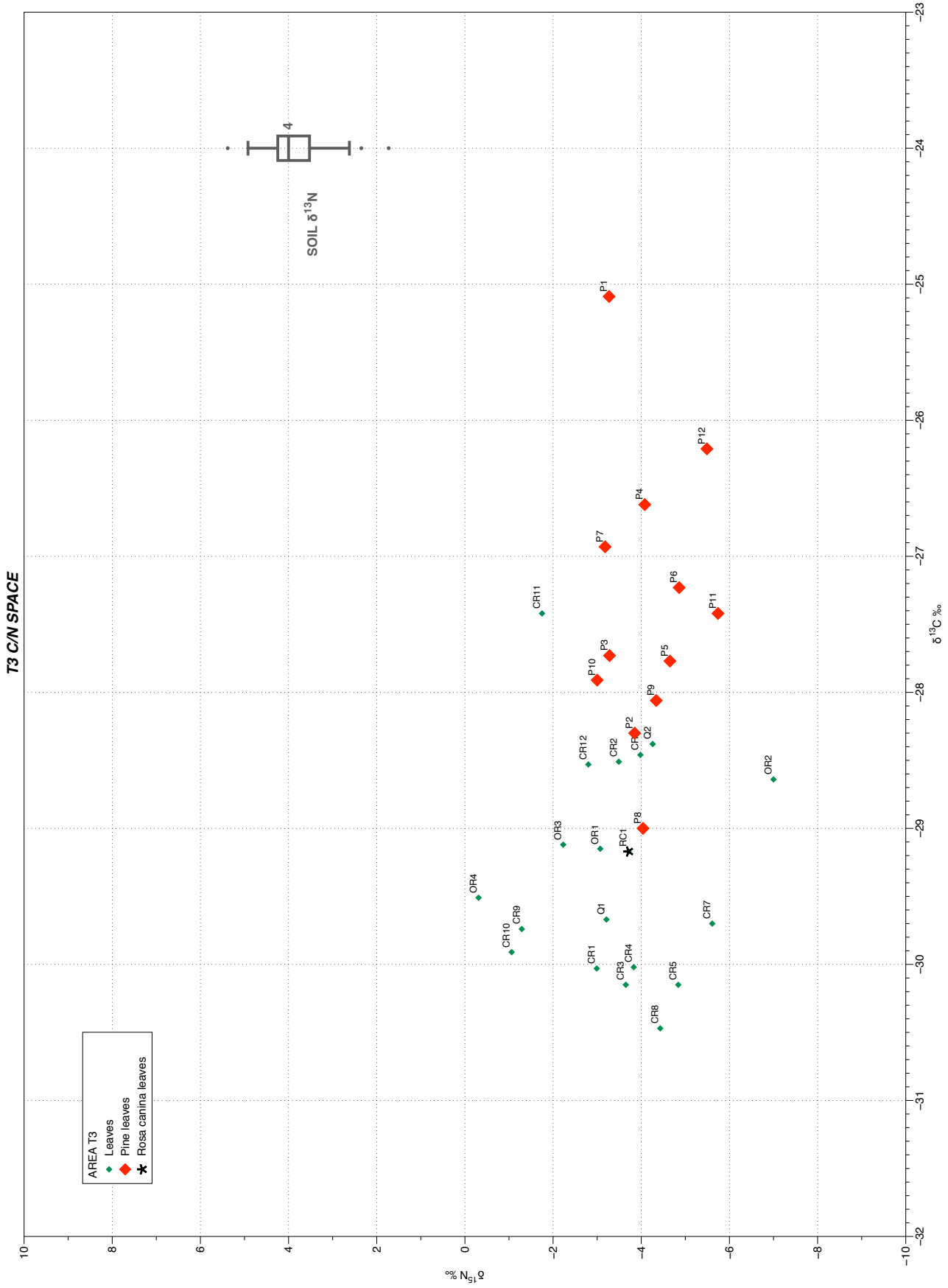
Figure 15: Evolution over time of the $\delta^{13}\text{C}$ vs $\delta^{15}\text{N}$ of truffles per site. The pine sites (T1, T2 and T8) show a marked ^{15}N enrichment. On the contrary, the areas without pines show a slight ^{15}N depletion corresponding, respectively, to the Symbiotic and Saprophytic regions of chapter 4. The ^{13}C variation is negligible and the overall variability of the mean $\delta^{13}\text{C}$ is less than 1‰, compared to the 1‰ of $\delta^{15}\text{N}$. The differences have been corrected according to the soil bias Δ_{SOIL} .



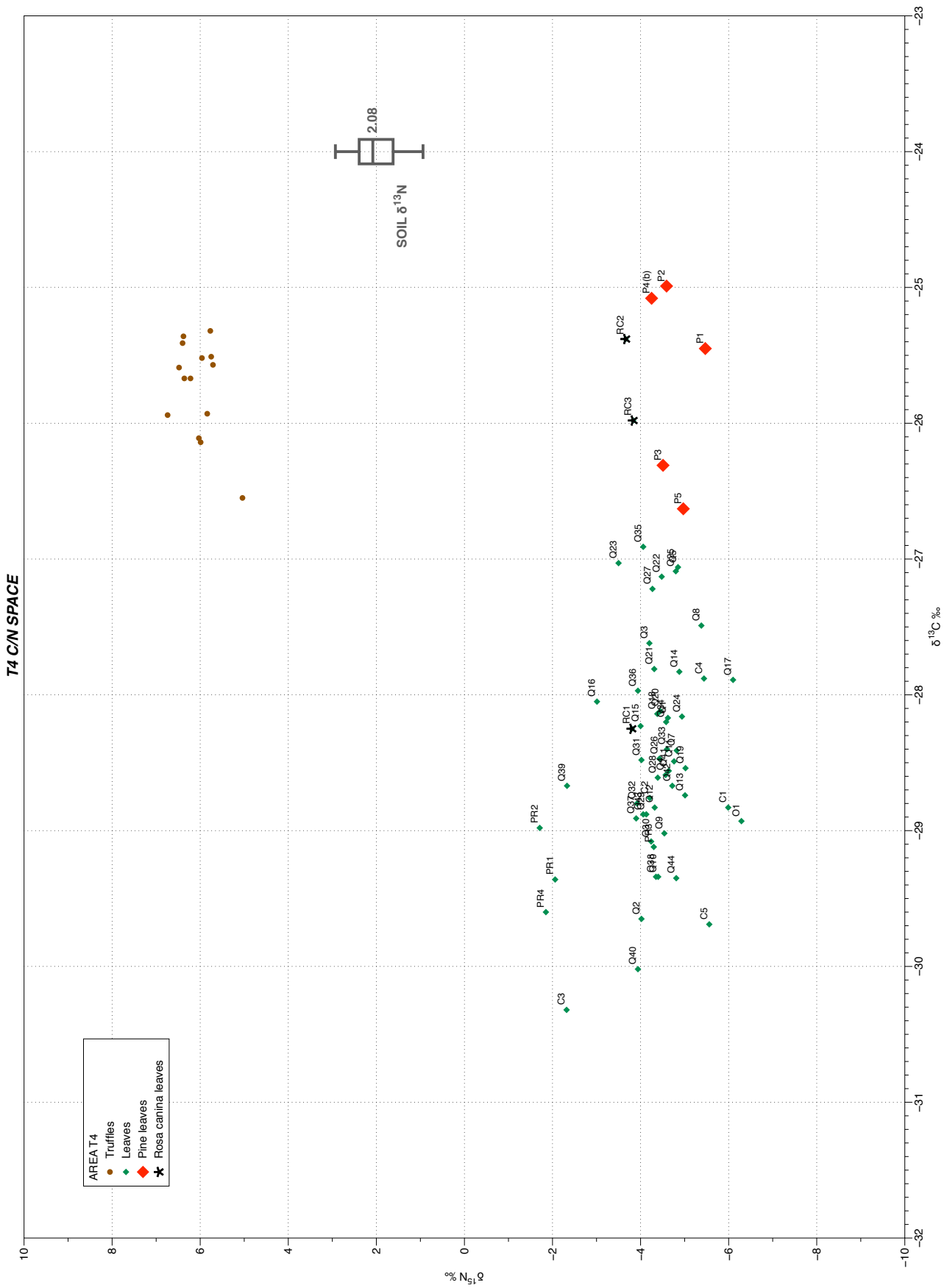
Site T1. Living pines. Symbiotic behaviour.



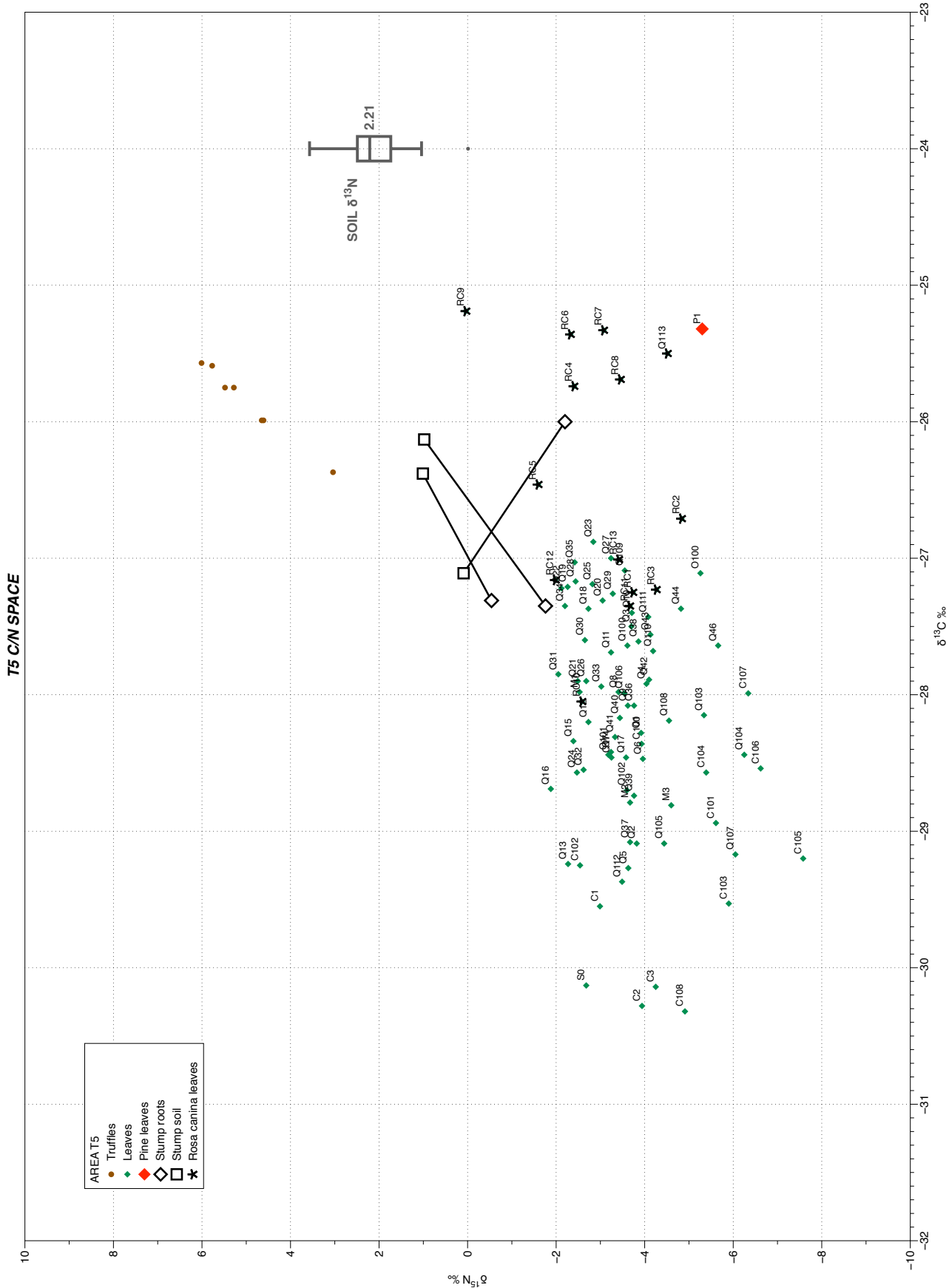
Site T2. Living pines. Symbiotic behaviour.



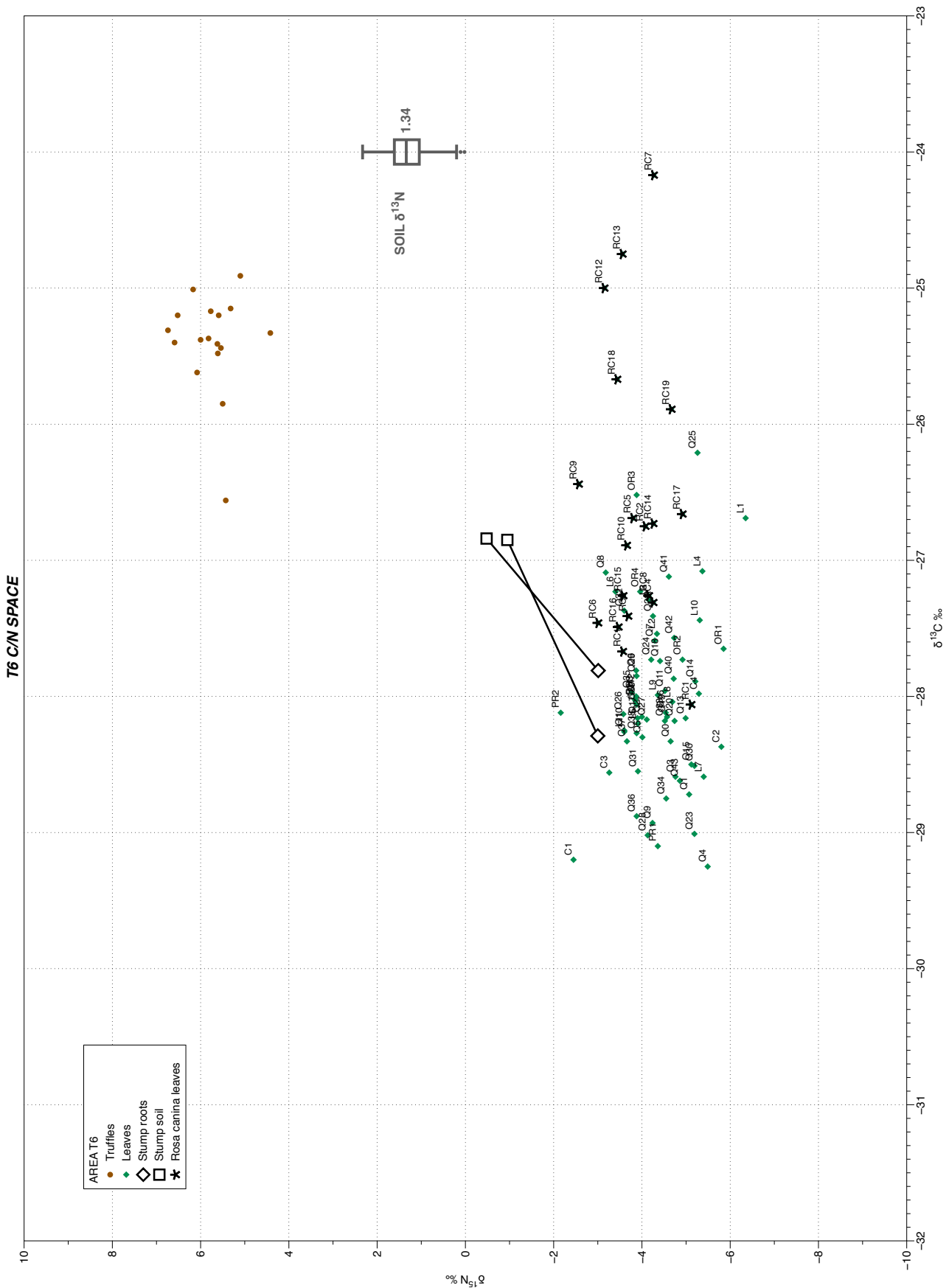
Site T3. Living pines, no known truffles production in year 2013 during the collection time.

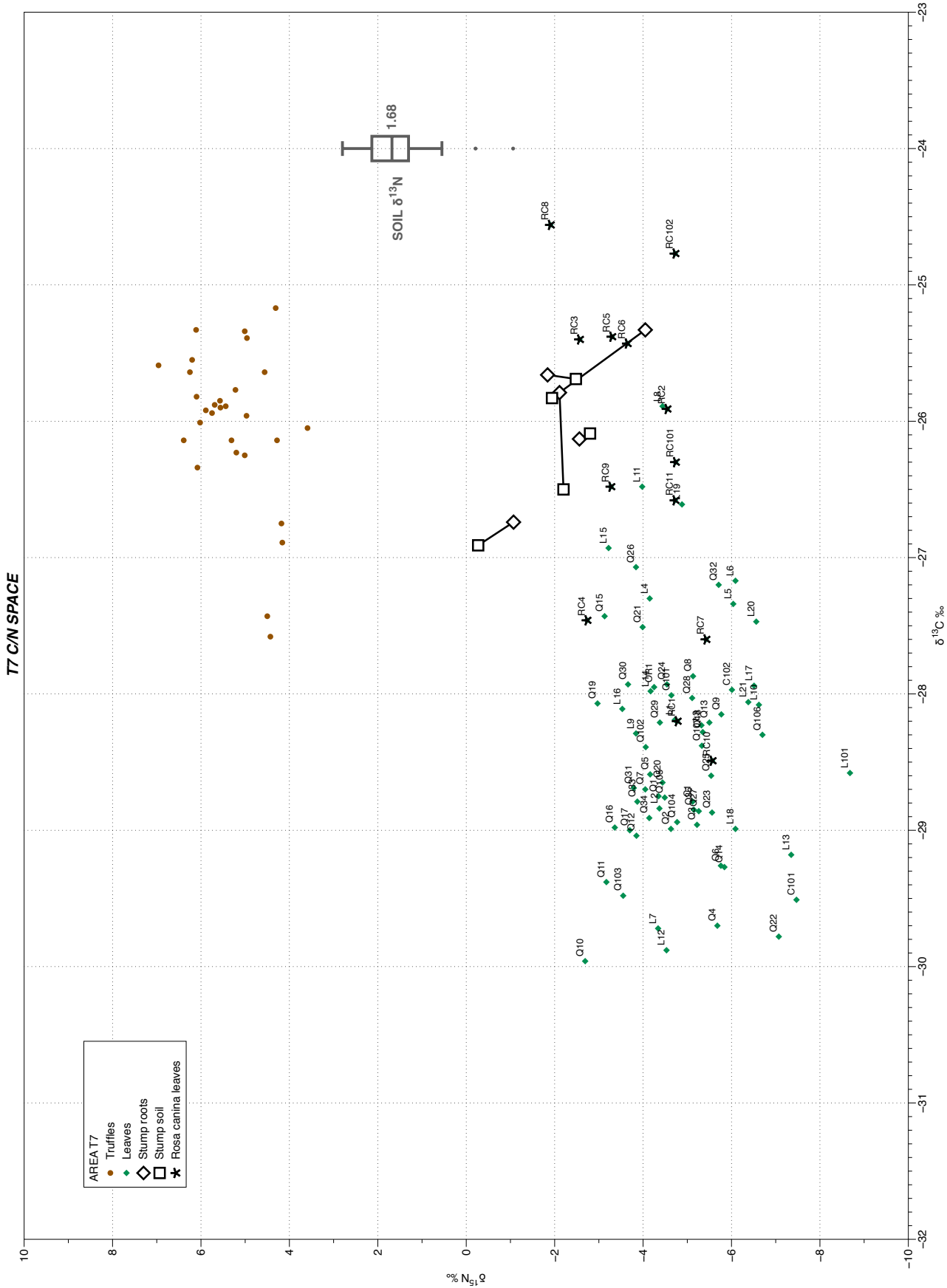


Site T4. A few living pines in the area. Uncertain behaviour.

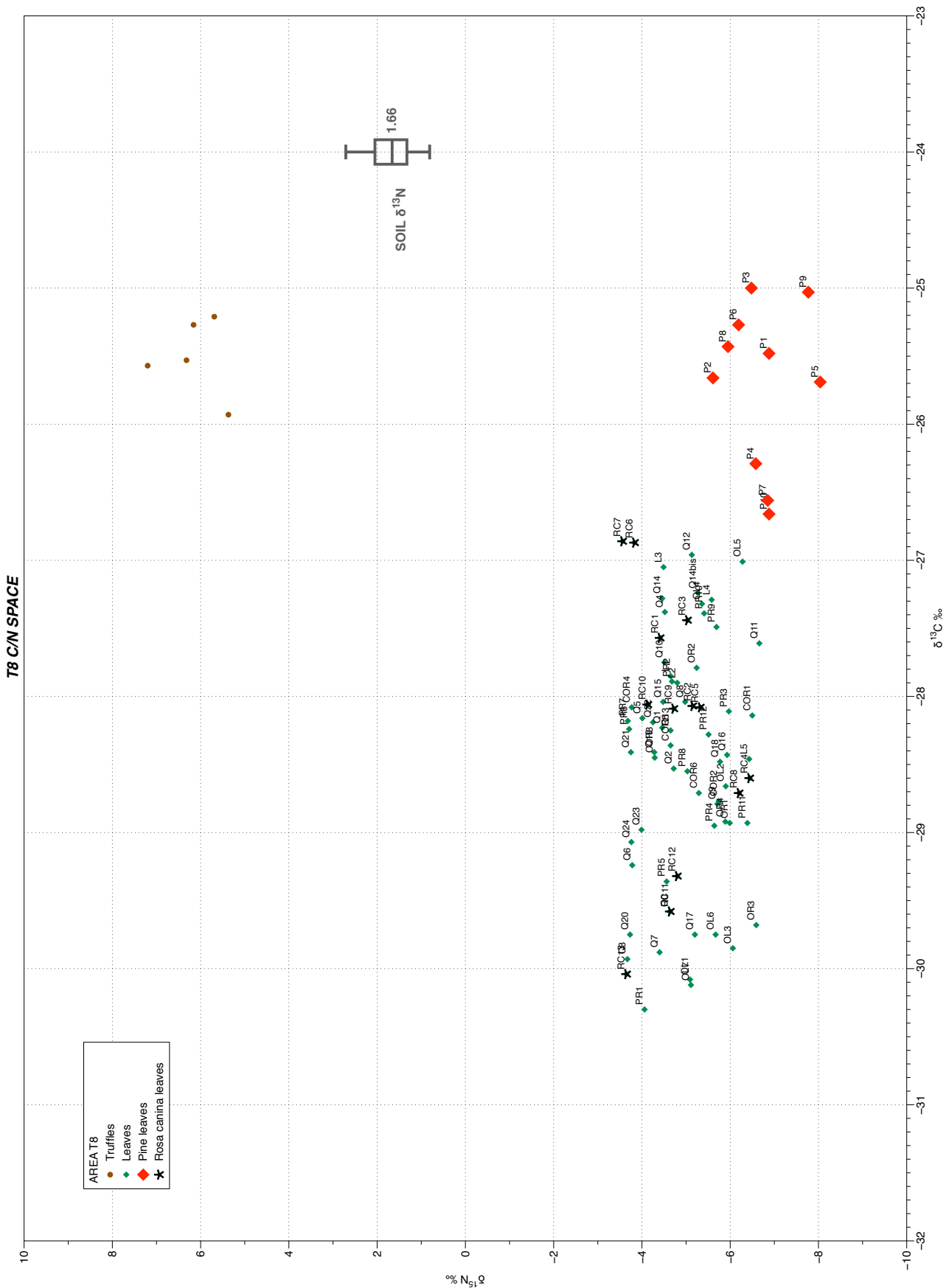


Site T5. A single pine close to the boundary of the area. Stumps found. Saprophytic behaviour.

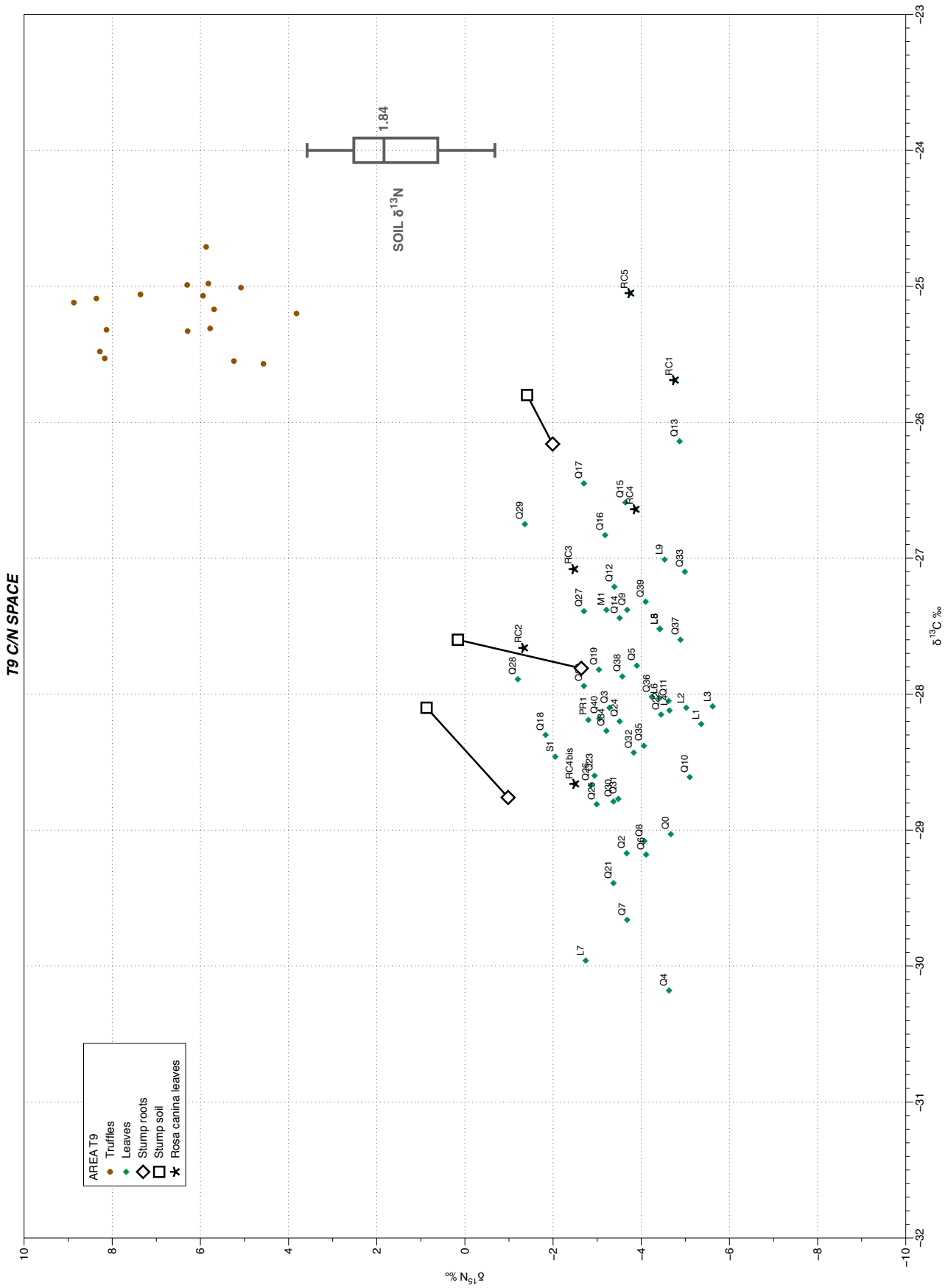




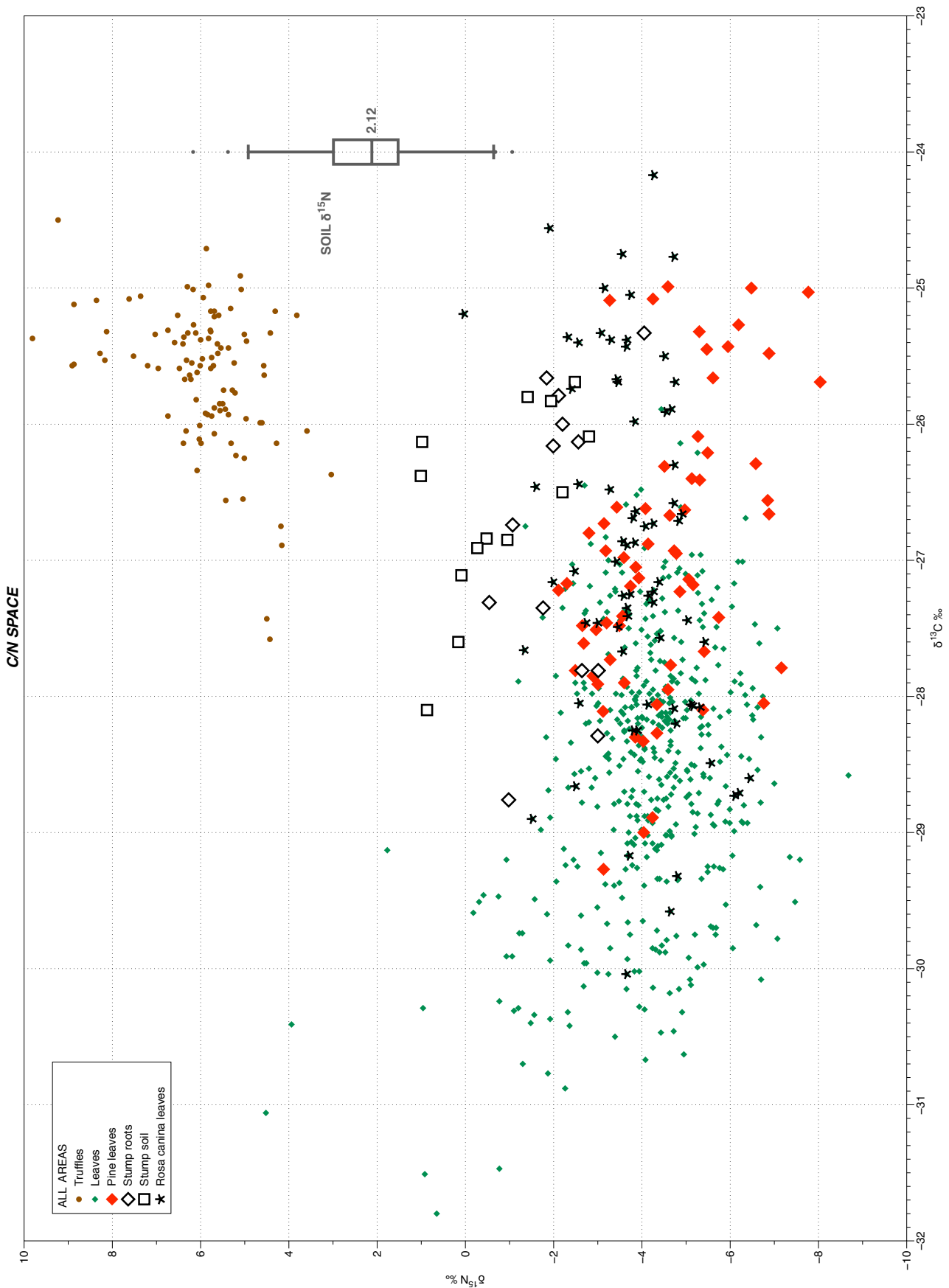
Site T7. Stumps found. Saprophytic behaviour.



Site T8. Living pines. Soil much different from T1, T2 and T3. Symbiotic behaviour.



Site T9. Stumps found. Saprophytic behaviour.



All sites' samples.

References

Literature

- [Agterberg 2014] Frits Agterberg - *Geomathematics: Theoretical Foundations, Applications and Future Developments*, Quantitative Geology and Geostatistics, vol.18 (2014). ISBN 978-3-319-06873-2. doi:10.1007/978-3-319-06874-9.
- [Angerosa et al. 1999] Angerosa F., Bréas O., Contento S., Guillou C., Reniero F., Sada E. - *Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils*, Journal of Agriculture and Food Chemistry, 47 (1999), 1013-1017.
- [Aquilina et al. 2006] Aquilina L., Ladouche B., Dorfliger N. - *Water storage and transfer in the epikarst of karstic systems during high flow periods*, Journal of Hydrology, 327 (34), 472-485 (2006). doi:10.1016/j.jhydrol.2005.11.054.
- [Barbour et al. 2005] Barbour M. M., Hunt J. E., Dungan R. J., Trunbull M. H., Brailsford G. W., Farquhar G. D., Whitehead D. - *Variation in the degree of coupling between $\delta^{13}\text{C}$ of phloem sap and ecosystem respiration in two mature Nothofagus forests*. New Phytologist, 166 (2005) 497-512. doi:10.1111/j.1469-8137.2005.01329.x.
- [Bence 1995] James R. Bence - *Analysis of Short Time Series: Correcting for Autocorrelation*, Ecology, Vol. 76, No. 2 (1995), 628-639. doi:10.2307/1941218.
- [Bivand et al. 2008] Roger S. Bivand, Edzer J. Pebesma, Virgilio Gómez-Rubio - *Applied Spatial Data Analysis with R*, Springer 2008. doi:10.1007/s11004-011-9340-y.
- [Borcard et al. 2011] Daniel Borcard, François Gillet, Pierre Legendre - *Numerical Ecology with R*, Springer 2011. ISBN 978-1-4419-7975-9. doi:10.1007/978-1-4419-7976-6.
- [Bogaert 1996] Patrick Bogaert - *Comparison of Kriging Techniques in a Space-Time Context*, Mathematical Geology, Vol.1 (1996). doi:10.1007/BF02273524.
- [Bowen and Revenaugh 2003] Bowen G. J. and Revenaugh J. - *Interpolating the isotopic composition of modern meteoric precipitation*, Water Resources Research, 39(10) 2003, 1299-1311. doi:10.1029/2003WR002086.
- [Bowen and Wilkinson 2002] Bowen G. J. and Wilkinson B. - *Spatial distribution of $\delta^{18}\text{O}$ in meteoric precipitation*, Geology, 30 (2002), 315-318. doi:10.1130/0091-7613.

- [Box et al. 2016] Box G.E.P., Jenkins G.M., Reinsel G., Ljung G. - *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2016. ISBN 978-0-470-27284-8.
- [Braun 1997] Gerald Braun - *The Use of Digital Methods in Assessing Forest Patterns in an Andean Environment: The Polylepis Example*, Mountain Research and Development Vol. 17 (1997) No.3, 253-262. doi:10.2307/3673852.
- [Brugnoli and Farquhar 2000] Enrico Brugnoli E, Graham D. Farquhar - *Photosynthetic Fractionation of Carbon Isotopes*, in *Advances in Photosynthesis*, vol.9 ch.17 99-434 (Richard C. Leegood, Thomas D. Sharkey, Susanne von Caemmerer eds.) Kluwer Academic Publishers 2000. ISBN 978-0-7923-6143-5. doi:10.1007/0-306-48137-5-17.
- [Camin et al. 2010a] Camin F., Larcher R., Nicolini G., Bontempo L., Bertoldi D., Perini M. - *Isotopic and elemental data for tracing the origin of European olive oils*. Food Chemistry, 58 (2010), 570-577. doi:10.1021/jf902814s.
- [Camin et al. 2010b] Camin F., Larcher R., Perini M., Bontempo L., Bertoldi D., Gagliano G., Versini G. - *Characterization of authentic Italian extra-virgin olive oils by stable isotopes ratios of C, O and H and mineral composition*. Food Chemistry, 118 (2010), 901-909.
- [Chiocchini et al. 2016] Chiocchini F, Portarena S, Ciolfi M, Brugnoli E, Lauteri M - *Isoscapes of carbon and oxygen stable isotope compositions in tracing authenticity and geographical origin of Italian extra-virgin olive oils*, Food Chemistry 202 (2016) 291-301. doi:10.1016/j.foodchem.2016.01.146.
- [Christakos 2000] Christakos G. - *Modern Spatiotemporal Geostatistics*, Studies in Mathematical Geology, Oxford University Press, New York, 2000. ISBN 978-0-486-48818-9.
- [Christakos et al. 2002] Christakos G., Bogaert P., Serre M.L. - *Temporal GIS: Advanced Functions for Field-based Applications*, Springer-Verlag, New York, 2002. ISBN 978-3540414766. doi:10.1007/978-3-642-56540-3.
- [Ciolfi et al. 2016a] Ciolfi M., Chiocchini F., Gavrichkova O., Pisanelli A., Portarena S., Scartazza A., Brugnoli E., Lauteri M. - *Timescape: a simple space-time interpolation geostatistical Algorithm*, 2016 EGU General Assembly, Geophysical Research Abstracts Vol. 18, EGU2016-15864, 2016. CC Attribution 3.0 License.
- [Ciolfi et al. 2016b] Ciolfi M., Chiocchini F., Russo G., Gavrichkova O., Pisanelli A., Portarena S., Scartazza A., Brugnoli E., Lauteri M. - *Spatial and spatio-temporal analysis of stable*

- isotopes in biogeochemical processes*, Proc. First Isotope ratio MS DAY, May 9-11 2016, F. Camin editor, ISBN 978-88-7843-046-4.
- [Cressie 1990] Noel A. C. Cressie - *Statistics for Spatial Data*, John Wiley & sons 1990.
- [Cressie and Wikle 2011] Noel Cressie, Christopher K. Wikle - *Statistics for Spatio-temporal Data*, Wiley Series in Probability and Statistics, John Wiley & sons 2011.
- [Cryer and Chan 2008] Jonathan D. Cryer, Kung-Sik Chan - *Time Series Analysis*, Springer Texts in Statistics, Springer 2008. ISBN 978-0-387-75958-6.
- [Dansgaard 1964] Dansgaard, W. - *Stable isotopes in precipitation*, Tellus, 16 (1964) 436-467. doi:10.1111/j.2153-3490.1964.tb00181.x.
- [Dawson et al. 2002] Todd E. Dawson, Stefania Mambelli, Agneta H. Plamboeck, Pamela H. Templer and Kevin P. Tu - *Stable Isotopes in Plant Ecology*, Annual Review of Ecology and Systematics, Vol. 33 (2002), 507-559.
- [Deckmyn et al. 2014] G. Deckmyn, A. Meyer, M.M. Smits, A. Ekblad, T. Grebenc, A. Komarov, H. Kraighe - *Simulating ectomycorrhizal fungi and their role in carbon and nitrogen cycling in forest ecosystems*, Canadian Journal of Forest Research, 2014, 44(6): 535-553, 10.1139/cjfr-2013-0496. doi:10.1139/cjfr-2013-0496.
- [Diggle and Ribeiro 2007] Peter J. Diggle, Paulo J. Ribeiro Jr. - *Model-based Geostatistics*, Springer Series in Statistics, Springer 2007. ISBN-13: 978-0-387-32907-9.
- [Fortin and Dale 2011] Marie-Josée Fortin, Mark Dale - *Spatial Analysis, a guide for the Ecologists*, Cambridge University Press, Cambridge 2005, eighth edition 2011.
- [Frankel 2012] Theodore Frankel - *The Geometry of Physics*, Cambridge University Press 2006, revised edition 2012. ISBN 978-1-107-60260-1.
- [Godbold et al 2006] Godbold D L, Hoosbeek M R, Lukac M, Cotrufo F, Janssens I A, Ceulemans R, Polle A, Velthorst E J, Scarascia-Mugnozza G, De Angelis P, Miglietta F, Peressotti A - *Mycorrhizal hyphal turnover as a dominant process for carbon input into soil organic matter*, Plant and Soil (2006) 281:1524. doi:10.1007/s11104-005-3701-6
- [Goodchild et al 1992] Michael Goodchild, Robert Haining, Stephen Wise - *Integrating GIS and spatial data analysis: problems and possibilities*, International Journal of Geographical Information Systems vol.6, 5, 1992. doi:10.1080/02693799208901923.

- [Goovaerts 1997] Goovaerts P. - *Geostatistics for natural resources evaluation*, Oxford University Press, Oxford 1997.
- [Govindarajulu et al. 2005] Manjula Govindarajulu, Philip E. Pfeffer, Hairu Jin, Jehad Abubaker, David D. Douds, James W. Allen, Heike Bücking, Peter J. Lammers, Yair Shachar-Hill - *Nitrogen transfer in the arbuscular mycorrhizal symbiosis*, Nature Letters Vol 435 (2005). doi:10.1038/nature03610.
- [Hamilton 1994] James D. Hamilton - *Time Series Analysis*, Princeton University Press 1994. ISBN 860-1300372280.
- [Henn and Chapela 2001] Matthew R. Henn, Ignacio H. Chapela - *Ecophysiology of $\delta^{13}C$ and $\delta^{15}N$ isotopic fractionation in forest fungi and the roots of the saprotrophic-mycorrhizal divide*, Oecologia (2001) 128:480-487. doi:10.1007/s004420100680.
- [Hijmans et al. 2005] Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G., Jarvis A. - *Very high resolution interpolated climate surfaces for global land areas*, International Journal of Climatology, 25 (2005), 1965-1978. doi:10.1002/joc.1276.
- [Hobbie et al. 2001] Erik A. Hobbie, Nancy S. Weber, James M. Trappe - *Mycorrhizal vs saprotrophic status of fungi: the isotopic evidence*, New Phytologist (2001) 150: 601-610.
- [Hobbie and Colpaert 2003] Erik A. Hobbie and Jan V. Colpaert - *Nitrogen availability and colonization by mycorrhizal fungi correlate with nitrogen isotope patterns in plants*, New Phytologist (2003) 157: 115-126. doi:10.1046/j.1469-8137.2003.00657.x.
- [Höegberg et al. 1999] Peter Höegberg, Mona. Höegberg, Maud E. Quist, Alf Ekblad, Torgny Näsholm - *Nitrogen isotope fractionation during nitrogen uptake by ectomycorrhizal and non-mycorrhizal Pinus sylvestris*, New Phytologist (1999), 142, 569-576.
- [Iacumin et al 2009] Iacumin P., Bernini L., Boschetti T. - *Climatic factors influencing the isotope composition of Italian olive oils and geographic characterisation*, Rapid Communications in Mass Spectrometry, 23 (2009), 448-454. doi:10.1002/rcm.3896.
- [Jassby and Powell 1990] Alan D. Jassby and Thomas M. Powell - *Detecting Changes in Ecological Time Series*, Ecology, Vol. 71, No. 6 (1990), 2044-2052. doi:10.2307/1938618.
- [Krasilnikov et al. 2008] Krasilnikov P., Carré, F. and Montanarella, L. (eds.) - *Soil geography and geostatistics: Concepts and Applications*, JRC Scientific & Technical Reports 44084, LB-NA-23290-EN-C, Luxembourg 2008. ISBN 978-92-79-08720-2 - ISSN 1018-5593.

- [Kyriakidis and Journel 1999] Phaedon C. Kyriakidis and Andr  G. Journel - *Geostatistical SpaceTime Models: A Review*, Mathematical Geology, Vol. 31, No. 6, 1999.
- [Lang et al 2013] Christa Lang, Reiner Finkeldey, Andrea Polle - *Spatial patterns of ectomycorrhizal assemblages in a monospecific forest in relation to host tree genotype*, Frontiers in Plant Science, 4 (2013), 448-454. doi:10.3389/fpls.2013.00103.
- [Lauteri et al 2004] Lauteri M., Pliura A., Monteverdi M. C., Brugnoli E., Villani F., Eriksson G. - *Genetic variation in carbon isotope discrimination in six European population of Castanea sativa Mill. originating from contrasting localities.*, Journal of Evolutionary Biology, 1286-1296 (2004). doi:10.1111/j.1420-9101.2004.00765.x.
- [Legendre and Legendre 2012] T. Pierre Legendre and Louis Legendre - *Numerical Ecology.*, 3rd edition, Elsevier 2012. ISBN 978-0-444-536868-0.
- [Liao 2005] T. Warren Liao - *Clustering of time series data: a survey.*, Pattern Recognition 38 (2005) 1857-1874. doi:0.1016/j.patcog.2005.01.025.
- [Longinelli and Selmo 2003] Longinelli A., Selmo E. - *Isotopic composition of precipitation in Italy: A first overall map.*, Journal of Hydrology, 270 (2003) 75-88.
- [Mayor et al 2009] Jordan R. Mayor, Edward A. G. Schuur, Terry W. Henkel - *Elucidating the nutritional dynamics of fungi using stable isotopes*, Ecology Letters, (2009) 12. 171-183. doi:10.1111/j.1461-0248.2008.01265.x.
- [Neteler et al. 2011] Markus Neteler, M. Hamish Bowman, Martin Landa, Markus Metz - *GRASS GIS: A multi-purpose open source GIS*, Environmental Modeling & Software. (2012) 124-130. doi:10.1016/j.envsoft.2011.11.014.
- [Nisbet et al. 2009] Nisbet R., Elder J., Miner G. - *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier 2009. ISBN 978-0-12-374765-5.
- [Oliver and Webster 2015] Margaret A. Oliver and Richard Webster - *Basic Steps in Geostatistics: The Variogram and Kriging*, Springer Briefs in Agriculture, Springer 2015.
- [Olivera et al 2011] Olivera A, Fischer C R, Bonet J A, Mart nez de Arag n J, Oliach D, Colinas J - *Weed management and irrigation are key treatments in emerging black truffle (Tuber melanosporum) cultivation*, New Forests (2011) 42:227-239.

- [Pedrycz and Chen 2013] Witold Pedrycz and Shyi-Ming Chen (eds) - *Time Series Analysis, Modeling and Applications, a Computational Intelligence Perspective*, Intelligent Systems Reference Library, Springer 2013. ISBN 978-3-642-33438-2. doi:10.1007/978-3-642-33439-9.
- [Peterson and Fry 1987] Bruce J. Peterson and Brian Fry - *Stable Isotopes in Ecosystem Studies*, Annual Review of Ecology and Systematics, Vol. 18 (1987), pp. 293-320.
- [Portarena et al 2014] Portarena S., Gavrichkova O., Lauteri M., Brugnoli E. - *Authentication and traceability of Italian extra-virgin olive oils by means of stable isotopes techniques*. Food Chemistry, 164 (2014), 12-16. doi:10.1016/j.foodchem.2014.04.115.
- [Portarena et al 2015] Portarena S., Farinelli D., Lauteri M., Famiani F., Esti M., Brugnoli, E. - *Stable isotope and fatty acid compositions of monovarietal olive oils: Implications of ripening stage and climate effects as determinants in traceability studies*. Food Control, 57 (2015), 129-135. doi:10.1016/j.foodcont.2015.03.052.
- [Renard et al. 2010] Philippe Renard, Hélène Demougeot-Renard, Roland Froidevaux (eds) - *Geostatistics for Environmental Applications*, Proceedings of the Seventh European Conference on Geostatistics for Environmental Applications. Springer 2005. ISBN-13 978-3-540-26533-7. doi:10.1007/978-90-481-2322-3.
- [Rossi et al. 1992] Richard E. Rossi, David J. Mulla, Andre G. Journel and Eldon H. Franz - *Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence*, Ecological Monographs, Vol.62 No.2 Jun 1992), pp.277-314, Wiley 1992.
- [Roßmann 2013] A. Roßmann - *New approaches for verifying the geographical origin of foods* in P. Brereton (ed) *New analytical approaches for verifying the origin of food*, Woodhead Publishing Series in Food Science, Technology and Nutrition. N.245 (2013).
- [Sarma 2009] D.O. Sarma - *Geostatistics with Applications In Earth Science*, the frontiers collection, Springer 2009. ISBN 978-1-4020-9379-1.
- [Schlosshauer 2008] Maximilian Schlosshauer - *Decoherence and the Quantum-to-Classical Transition*, the frontiers collection, Springer 2008. ISBN 978-540-35773-5.
- [Sethna 2006] James P. Sethna - *Statistical Mechanics: Entropy, Order Parameters and Complexity*, Oxford Master Series in Statistical, Computational and Mathematical Physics, Clarendon Press - Oxford University Press 2006.

- [Simard et al. 1997] Suzanne W. Simard, David A. Perry, Melanie D. Jones, David D. Myrold, Daniel M. Durall, Randy Molinak - *Net transfer of carbon between ectomycorrhizal tree species in the field*, Letters to Nature Vol. 388 (1997) 279-581.
- [Simpson 1979] Beryl B. Simpson - *A Revision of the Genus Polylepis*, Smithsonian Contributions to Botany, Smithsonian Institution Press, Washington 1979.
- [Slater et al. 2001] Christine Slater, Tom Preston and Lawrence T. Weaver - *Stable isotopes and the international system of units*, Rapid Commun. Mass Spectrom. 2001; 15: 1270-1273. doi:10.1002/rcm.328.
- [Steiniger and Hay 2009] Stefan Steiniger, Geoffrey J. Hay - *Free and open source geographic information tools for landscape ecology*, Ecological Informatics 4 (2009) 183-195. doi:10.1016/j.ecoinf.2009.07.004.
- [Steiniger and Hunter 2013] Stefan Steiniger, Andrew J.S. Hunter - *The 2012 free and open source GIS software map. A guide to facilitate research, development and adoption*, Computers, Environment and Urban Systems 39 (2013) 136-150. doi:10.1016/j.compenvurbsys.2012.10.003.
- [Stevens 2009] M. Henry H. Stevens - *A Primer of Ecology with R*, Use R! series, Springer 2009. ISBN 978-0-387-89881-0. doi:10.1007/978-0-387-89882-7.
- [Szeckeres 2006] Peter Szeckeres - *Modern Mathematical Physics: Groups, Hilbert Spaces and Differential Geometry*, Cambridge University Press 2006. ISBN 0-521-82960-7.
- [Turchin and Taylor 1992] Peter Turchin and Andrew D. Taylor - *Complex Dynamics in Ecological Time Series*, Ecology, Vol. 73, No. 1 (1992), 289-305. doi:0.2307/1938740.
- [Urbano and Cagnacci 2014] Ferdinando Urbano and Francesca Cagnacci (eds.) - *Spatial Database for GPS Wildlife Tracking Data*, Springer 2014. ISBN 978-3-319-03742-4. doi:10.1007/978-3-319-03743-1.
- [van der Veer 2013] van der Veer G. - *Development and application of geospatial models for verifying the geographical origin of food*, In Brereton P (ed). *New analytical approaches for verifying the origin of food* pp. 6080). Woodhead Publishing Series in Food Science, Technology and Nutrition. No. 245 (2013) 60-80. ISBN 978-0-85709-274-8.
- [Visser 2006] Darcy R. Visser - *GPS measurement error and resource selection functions in a fragmented landscape*, Ecography vol.39, 3, 2006. doi:10.1111/j.0906-7590.2006.04648.x.

- [Wagner and Fortin 2005] Helene H. Wagner and Marie-José Fortin - *Spatial Analysis of Landscape: Concepts and Statistics*, Ecology, 86(8), 2005, 1975-1987.
- [West et al. 2006] West J. B., Bowen G. J., Cerling T. E., Ehleringer J. R. - *GStable isotopes as one of nature's ecological recorders. Trends in Ecology and Evolution*, Trends in Ecology and Evolution, 21(7), 2006, 408-414. doi:10.1016/j.tree.2006.04.002.
- [West et al. 2010] Jason B. West, Gabriel J. Bowen, Todd E. Dawson, Kevin P. Tu eds. - *Isoscapes, Understanding Movement, Pattern and Process on Earth Through Isotope Mapping*, Springer 2010. ISBN 978-90-481-3353-6. doi:10.1007/978-90-481-3354-3.
- [Werner and Schmidt 2002] Roland A. Werner, Hanns-Ludwig Schmidt (review) - *The in vivo nitrogen isotope discrimination among organic plant compounds*, Phytochemistry 61 (2002) 465-484. doi:10.1016/S0031-9422(02)00204-2.
- [Woodhouse 2003] Nicholas M. J. Woodhouse - *Special Relativity*, Springer Undergraduate Mathematics Series, Springer 2003. ISBN 1-85233-426-6.
- [Zuur et al. 2007] Alan F. Zuur, Elena N. Ieno, Graham M. Smith - *Analysing Ecological Data* - Springer 2007. ISBN 0-387-45967-7.

Software and Network Resources

- [ARCGIS] ArcGIS™ is a registered software of Environmental Systems Research Institute (ESRI) - Redlands (CA) USA: <http://www.esri.com/software/arcgis>.
- [APACHE] The Apache Commons Mathematics Library: <http://commons.apache.org/proper/commons-math/>
- [ECMA] An ECMAScript reference: <http://www.ecma-international.org/publications/standards/Ecma-262.htm>
- [ECUAFLUX] Ecuafux project: <http://www.ub.edu/ecologia/ecuaflux/>
- [GEO TIFF] Niles Ritter, Mike Ruth - *GeoTIFF Format Specification*: <http://www.pubdoc.org/fileformat/rasterimage/tiff/geotiff.pdf>
- [GNU] GNU General Public License: <http://www.gnu.org/licenses/gpl-3.0.en.html>.

- [HIBERNATE] Hibernate ORM Idiomatic persistence for Java and relational databases:
<http://hibernate.org/orm/>.
- [HYNDMAN] Rob J Hyndman (Rob.Hyndman@monash.edu) CRAN Task *Time Series Analysis*
documentation: <https://cran.r-project.org/web/views/TimeSeries.html>.
- [ISOMAP] Isomap: Isoscape modeling, analysis and prediction:
<http://isomap.rcac.purdue.edu:8080/gridsphere/gridsphere>
- [JAVA] Oracle™ Java software main page: <http://www.oracle.com/technetwork/indexes/downloads/>
- [MYSQL] MySQL™ Community Edition: <https://www.mysql.com/products/community/>
- [NATURALEARTHDATA] : Natural Earth Data free vector and raster worldwide datasets:
<http://www.naturalearthdata.com>
- [ORACLE] Oracle™ Database Express Edition 11g: <http://www.oracle.com/technetwork/database/database-technologies/express-edition/downloads/index.html>
- [PCN] Portale Cartografico Nazionale: <http://www.pcn.minambiente.it/>; WMS services:
<http://www.pcn.minambiente.it/GN/en/accesso-ai-servizi-eng/wms-display-service>
- [QGIS] QGIS open source Geographic Information System: <http://www.qgis.org>
- [R] The R Project for Statistical Computing: <https://www.r-project.org>
- [RSTUDIO] RStudio is an open source user-friendly way to use R: <https://www.rstudio.com>
- [TIMESCAPEGLOBAL] TimescapeGlobal geostatistical package, released under GNU/GPLv3
license: <https://sourceforge.net/projects/timescapeglobal/>
- [WORLDCLIM] WorldClim - Global Climate Data: <http://www.worldclim.org>

Institutions

- [CBA] Centro de biologia Ambiental de a Faculdade de Ciencias de l'Universidade de Lisboa, Edificio C2, 5 Piso, Sala 2.5.46, Campo Grande 1749-016 Lisboa, Portugal.
<http://cba.fc.ul.pt>

- [GNIP] Global Network of Isotopes in Precipitation (GNIP) of the International Atomic Energy Agency (IAEA). http://www-naweb.iaea.org/napc/ih/IHS_resources_gnip.html
- [IBAF] Istituto di Biologia Agroambientale e Forestale del Consiglio Nazionale delle Ricerche, viale Marconi 2, 05010 Porano, Italy. <http://www.ibaf.cnr.it/en/>
- [SPATIAL] The SPATIAL (SPAtio-Temporal Isotope Analytics Lab) group, Department of Geology & Geophysics of the University of Utah, Frederick Albert Sutton bld. 115 S 1460 E Salt Lake City (UT) USA. <http://wateriso.utah.edu/spatial/>
- [UNAPROL] UANPROL Consorzio Olivicolo Italiano, Via XXIV Maggio 43, 00187 Roma. <http://www.unaprol.it>

Index

- δ notation, 54
- $g_{\mu\nu}$, 81
- Algorithm, 14, 21, 62, 80, 88
 - causal estimation, 89
 - jackknifing, 90
 - main loop, 88
 - model evaluation, 89
 - parameters
 - ensemble estimate, 92
 - recursive estimate, 91
 - preliminary steps, 86, 88
- Behaviour
 - saprophytic, 52, 60
 - symbiotic, 49, 54, 60
- Carbon flow, 52
- Causal cone, 19, 23, 83, 85, 91, 92
- Causality, 18
- Christoffel symbol, 81
- Curved space, 81
- Database, 24, 93
 - algorithm, 80
 - ER model, 93
 - logical structure, 96
 - tables, 93
- Distance, 16, 80, 81
 - space, 81, 82
 - time, 80, 82
- Ecuaflex, 73
- Elliptic integral, 81
- Ensemble, 90
- Ensemble estimation, 92
- EVOO, 63, 64
 - ^{13}C isoscape, 71
 - ^{18}O isoscape, 70
 - confidence intervals, 67
- Flat space, 82
- Food fraud, 63, 72
- Geodesic, 17
- Geodesic arc, 82
- Geodesic distance, 81
- Georeferencing, 53
- Geostatistics, 9, 12
- GPS, 3, 53, 75
- GUI, 27, 32, 40, 45
- IDW, 9, 14, 22, 87
- IRMS, 54, 64
- Isoscape, 49, 64
 - ^{13}C
 - EVOO, 66, 71
 - leaves, 56
 - truffles, 58
 - ^{15}N
 - leaves, 56
 - soil, 58
 - truffles, 58
 - ^{18}O
 - EVOO, 66, 70
 - confidence intervals, 69
- Jackknifing, 90
- Java, 93

- Kriging, 9, 22, 55, 66
 - ancillary, 93, 94
 - ancillary_source, 93, 94
 - model, 93, 94
 - model_source, 93, 95
 - source, 93, 94
 - voxel, 94, 95
- Lattice
 - pixel, 14
 - voxel, 16
- Metric, 81, 83
 - diamond, 83
 - Euclidean, 83
 - square, 83
- Metric tensor, 81
- Model
 - definition, 86
- Mycorrhiza, 49, 60
 - statistics, 61, 110
- MySQL, 94
- Nitrogen flow, 52
- OLS, 66
- Oracle, 96
- Polar coordinates, 81, 82
- Projected coordinates, 82
- Recursive estimation, 91
- Regression, 66
- Sampling atlas, 97
- Sampling planning, 73
- Simple Kriging, 66
- Software, 25
- Spatiotemporal modeling, 12
- Spacetime, 11, 12
- Spatial modeling, 9
- Symbiosis, 52
- Table