



HAL
open science

Dendro-assistance à la détermination de la provenance géographique des bois archéologiques

Georges Lambert

► **To cite this version:**

Georges Lambert. Dendro-assistance à la détermination de la provenance géographique des bois archéologiques. ARCADE Approche diachronique et Regards croisés : Archéologie, Dendrochronologie et Environnement, 2014, Aix-en-Provence, France. pp.33-49. hal-02143048

HAL Id: hal-02143048

<https://hal.science/hal-02143048v1>

Submitted on 5 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dendro-assistance à la détermination de la provenance géographique des bois archéologiques

*Georges Lambert**

Avec les données de:
Jean-Luc Dupouey (INRA), Patrick Gassmann (Musée Laténium, Neuchâtel, Suisse),
Olivier Girardclos (CNRS), Frédéric Guibal (CNRS), Patrick Hoffsummer (ULg, Liège),
Georges Lambert et Catherine Lavier (CNRS), François Lebourgeois (INRA), Yvonne Trenard (CTB, Paris)

Résumé

L'objectif est de montrer comment les corrélations dendrochronologiques peuvent aider à la détermination des sources anciennes d'approvisionnement en bois. On utilise ici des matrices de corrélations croisées du chêne. Ces matrices de corrélations croisées de bois déjà tous inter-datés ou inter-corrélés seront rangées par une classification automatique. La méthode utilisée est une classification hiérarchique ascendante qui permet de regrouper les bois en familles. L'algorithme commence par rassembler les séries les plus « proches » dans des familles les plus homogènes possible. Puis, progressivement, par étapes successives, il étend les familles trouvées aux plus proches séries restantes jusqu'à ne constituer qu'une seule famille avec l'ensemble de la population à traiter. Le résultat est communiqué dans un graphique appelé « arbre hiérarchique ascendant » ou « arbre typologique ». Les membres de chaque famille sont reliés par des « branches » (ici horizontales) dont la longueur est *inversement* proportionnelle à la corrélation. Plus la branche est courte, plus les croissances sont semblables, plus la « famille » est homogène. L'objectif est donc de tester la pertinence de ce type de calcul sur des arbres carottés vivants parfaitement localisés. Une première série de tests porte sur 129 arbres prélevés dans 16 stations dispersées dans une même forêt sur une vingtaine de kilomètres (Forêt de Chaux, Jura). Une deuxième série de tests portera sur 155 arbres, extraits d'un lot de plus d'un demi-millier d'arbres, issus de 32 forêts réparties en France sur un large transect Est-Ouest d'environ 800 km de long. Une forêt du canton Suisse de Neuchâtel, et une autre de la province belge de Liège ont été intégrées à ce lot.

Les résultats sont intéressants mais non miraculeux. Sous les contraintes imposées à la démonstration – notamment la nécessité de ne pas travailler sur des séries typologiquement isolées –, l'algorithme sépare avec pertinence les familles d'arbres provenant du même lieu. Mais il ne livre pas de critère fiable qui permettrait d'estimer la distance entre ces familles. L'archéologue, l'historien ou l'expert du monument ancien ont donc loisir de mobiliser toute leur culture professionnelle pour appuyer telle ou telle interprétation des résultats affichés. Tous les calculs ont été effectués dans le logiciel expérimental Dendron-IV (Lambert 2006).

Mots-clés : dendrochronologie, modélisation mathématique, dendro-provenance, France, chêne.

Abstract

We want to show how to help the identification of geographical sources of archaeological or ancient woods using the dating correlations. We use here cross correlation matrices between ring width series. These matrices are ordered by an automatic classification called “ascendant hierarchical classification”. The algorithm gathers first the closer series in elementary “families” as more homogeneous as possible. Then, it extends the found families to the next closer series not yet included and so on, step by step, the algorithm builds a “tree” of families until the whole population is completely integrated in one big family. The result is visible on a graphic called “ascendant hierarchical tree”. Members of each family are connected by “branches” which of the length is *inversely* proportional to the correlation. More the “branch” is short (here the branches are horizontal), more the correlation is big, more the resemblance is great, more the family is homogeneous. Then, the aim consists to test the pertinence of this kind of computation against living trees (oak trees) perfectly located. The first part of the paper treats 16 places of one forest (Chaux Forest, Jura, France) spread along an about 20 km long transect. The second part treats 32 places of forests spread along a

* Collaborateur de l'université de Liège (ULg, B), chercheur honoraire du CNRS.

French transect West-East of about 800 km. Elements from the Belgian province of Liège and other ones from the Swiss canton of Neuchâtel were included in the computed set.

Results are interesting but not miraculous. Under forced constraints – especially do not use isolated typological series –, the algorithm separates clearly enough trees that come from the same location. But it does not deliver reliable criterion for the distance estimation between the places. So the professional culture and knowledge of our colleagues who are not dendrochronologists may completely work to give sense to such results. Computations made by the experimental software Dendron-IV (Lambert 2006).

Keywords: dendrochronology, mathematics models, dendro-provenance, France, Oak tree.

INTRODUCTION

La dendrochronologie permet de dater les bois trouvés dans les fouilles archéologiques ou les monuments anciens. La datation est basée sur la comparaison de la variation de la largeur des cernes annuels d'arbres différents, les arbres ou bois datés communiquant la date aux bois non datés. Les croissances d'arbres poussant au même endroit sont généralement très semblables. Il va de soi que les autres conditions environnementales, autres que l'éloignement jouent un rôle – ô combien non négligeable – dans la variation de la production du cerne : texture et hydrologie du sol, compétition, etc. Mais, en dendrochronologie archéologique, ces derniers facteurs sont quasiment

inconnus. Les chercheurs s'appliquent cependant à lever des critères, les moins imprécis possible, qui permettraient de définir un périmètre probable de la zone d'approvisionnement en bois. La dendrochronologie étant comparative, cette démarche est en arrière-plan de tout travail de datation¹. Cependant, il n'est pas rare que les bois d'un même bâtiment, proviennent de lieux différents et la question se complique quand les bâtisseurs s'approvisionnent aussi dans des dépôts du commerce à moyenne et longue distance, notamment du commerce par voie d'eau. Pendant la procédure de datation, le dendrochronologue identifie fréquemment, d'après la typologie variable de ses séries dendrochronologiques, des provenances différentes plus ou moins contemporaines. À défaut de publications

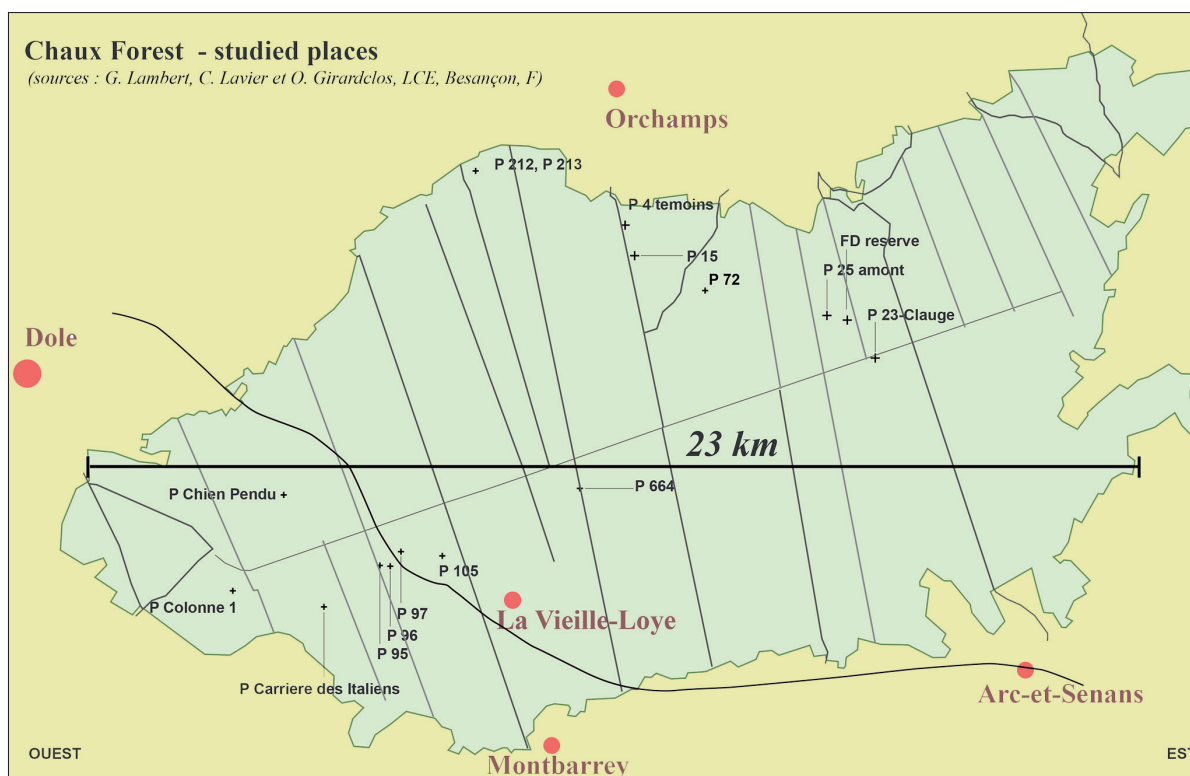


Fig. 1 – Étendue de la Forêt de Chaux, au nord du Jura Français, zone de prélèvements test sur du chêne. Les arbres ont été carottés par groupes en 16 points répartis dans les vingt mille hectares de cette forêt majoritairement domaniale. Cet espace sert ici de base expérimentale pour initier des calculs de dendro-provenance. Les stations sont diversement espacées sur un cailloutis argileux de l'époque tertiaire (système Aar-Doubs) percé ici et là de pointes du soubassement calcaire.

1. Voir par exemple Lambert *et al.* 2008.

French forests used in this study
The computing period is 1910 - 1980 AD (CE)

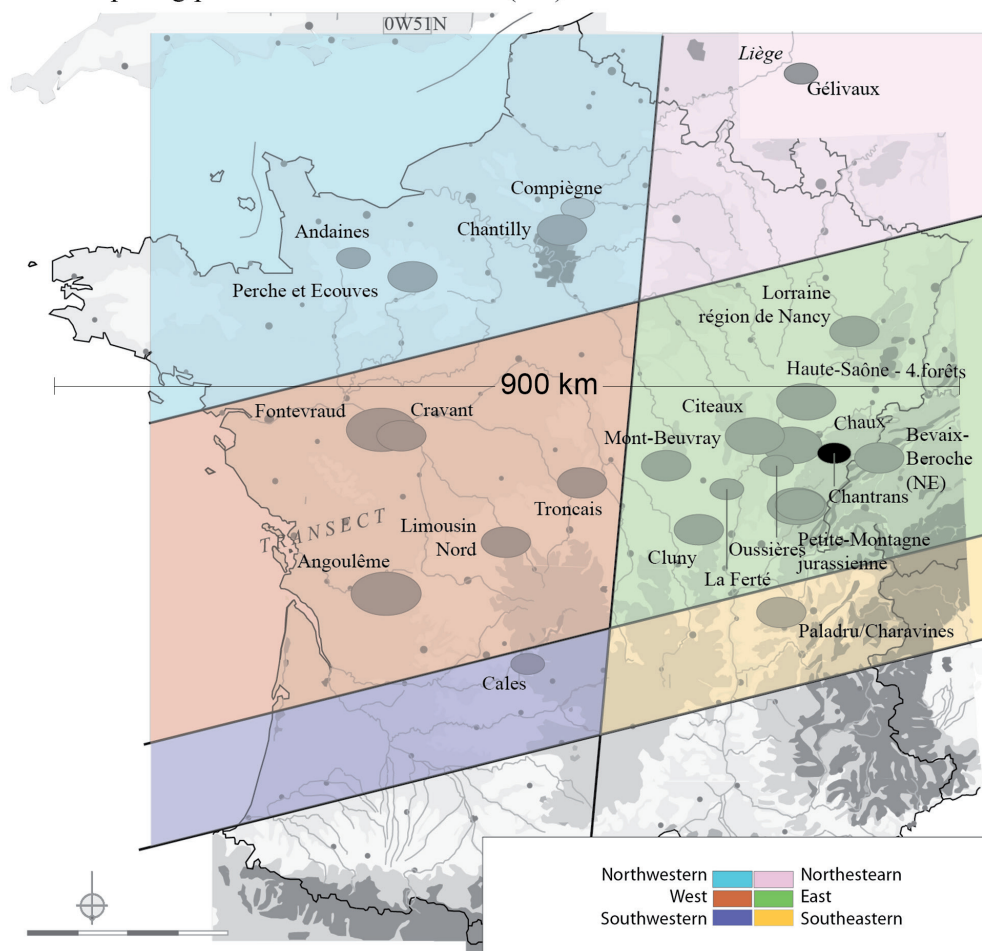


Fig. 2 – Stations forestières ou boisées échantillonnées en France Tempérée, Belgique du sud et Suisse occidentale. Ici, seul le chêne est concerné. Les laboratoires du CNRS (Besançon), de l'INRA (Nancy-Champenoux) ainsi que les laboratoires de l'université de Liège (ULg, Belgique) et du Musée Laténium de Neuchâtel (Suisse), ont fourni la plupart des échantillons. Les vents dominants atlantiques d'Ouest en Est constituent le principal facteur d'homogénéisation du signal dendrochronologique. La discrimination territoriale doit donc se moduler sur ce vecteur qui évolue d'une humidité atlantique étalée dans l'année vers des sécheresses aléatoires continentales.

de synthèse, les rapports privés des laboratoires de datation sont chargés de remarques traitant de cette question de l'origine des bois. Le traitement de cette question, la dendro-provenance, met en œuvre des faisceaux de considérations contextuelles : économiques, historiques, accidentelles (accidents météorologiques connus par exemple), etc.

En nous basant sur l'analyse dendrochronologique d'arbres vivants, carottés sur pied et par groupes (stations forestières), donc précisément localisés, nous évaluons dans ce papier la légitimité d'entreprendre une démarche de dendro-provenance à partir du seul signal dendrochronologique.

Ce texte est composé de deux parties qui suivront la présentation de la méthode. La première partie traite la question à l'échelle locale, échelle ordinaire d'un marché du bois d'une

bourgade au XV^e siècle par exemple. La forêt domaniale de Chauv, à l'est de Dole (Jura) nous offre un espace de 20 mille hectares dans lequel 129 arbres ont été prélevés par groupes dans 16 stations forestières, réparties sur près de 20 km (fig. 1). Visant à traiter des ensembles dendrochronologiques plus vastes, nous allons tirer de cette première partie, une technique d'échantillonnage des séries pour traiter 32 stations réparties sur l'ensemble de la France tempérée, la Belgique ardennaise et la Suisse occidentale (population totale 555 arbres, fig. 2). De la leçon tirée de la première partie, un échantillon de cette population (2 à 5 arbres par station, soit un échantillon de 155 arbres) est traité dans la seconde partie. Commençons par un rappel sur la méthode d'analyse que nous avons largement décrite ailleurs².

2. Lambert 2006.

MÉTHODE

La datation étant acquise (il s'agit d'arbres vivants connus sur pieds et carottés à une date précise), les calculs porteront sur une période fixe de 70 ans, 1910-1980 et les séries sont calibrées par la technique du corridor (fig. 3). Le coefficient de corrélation ou sa probabilité estimée par le test de Student seront utilisés comme «distance» entre les séries. La distance typologique entre les séries n'est pas la distance kilométrique mais une distance graphique qui permet d'évaluer d'un coup d'œil l'homogénéité du signal d'un groupe sur un graphique: ici, soit dans une matrice de corrélations croisées, soit dans un «arbre hiérarchique».

La matrice de corrélations croisées est un tableau carré (autant de lignes que de colonnes) des corrélations de toutes les séries sur toutes les autres. L'ordre des séries est le même en ligne qu'en colonne et la matrice se lit de la

même façon de gauche à droite que de haut en bas. Elle est symétrique: par exemple, la corrélation entre la série 4 et la série 7 se trouve portée deux fois: une fois dans la case ou colonne 4 de la ligne 7 et une deuxième fois dans la case ou colonne 7 de la ligne 4. La diagonale 1 de la matrice qui représente la corrélation de chaque série sur elle-même, est vide. Les calculs n'utilisent qu'une moitié de la matrice, située d'une part ou de l'autre de la diagonale 1. La présentation complète des matrices utilisée ici (sauf la dernière) n'est qu'un confort visuel ou une habitude, la moitié située en dessous de la diagonale suffit. La lecture des coefficients de corrélation précis ou des probabilités n'étant pas utile ici, les nombres ont été remplacés par des nuances de gris (10 paliers), le noir parfait représente la corrélation maximale (corrélation 1 ou probabilité 1), le blanc parfait représente la corrélation zéro ou la probabilité zéro. Le niveau de corrélation général étant assez bon, même à des distances de plusieurs centaines de kilomètres,

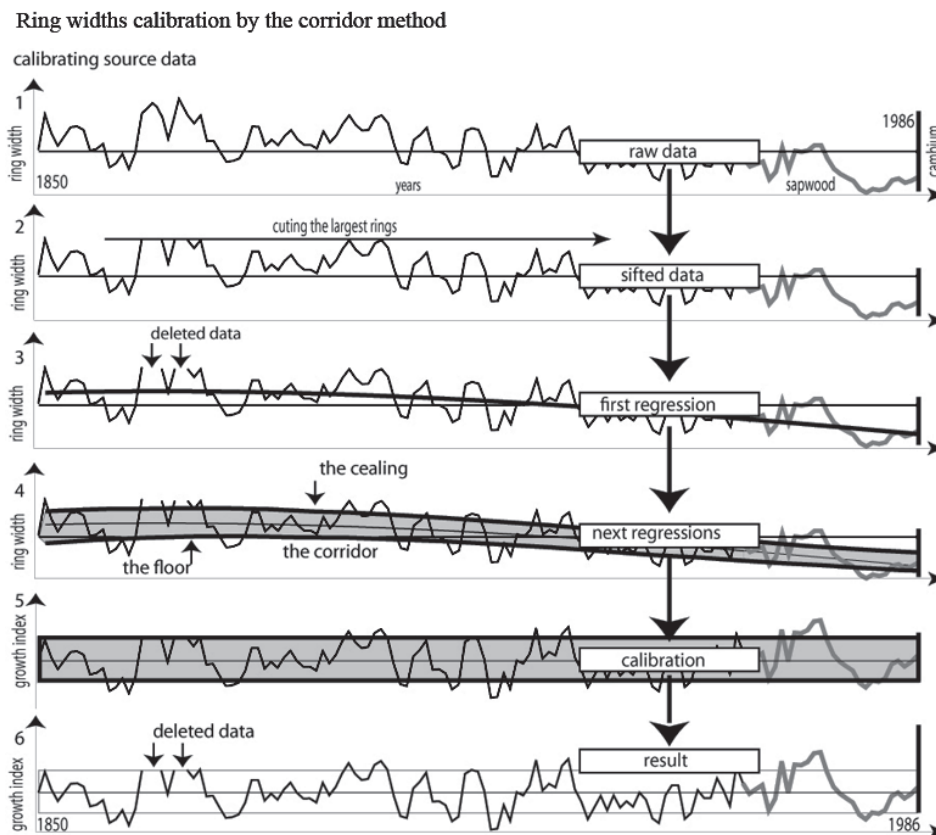


Fig. 3 – Méthode de calibration du signal dendrochronologique brut. Les largeurs de cernes sont incluses dans un «corridor» défini par 3 régressions polynomiales du même degré (en général du 3^e degré). Une première régression est calculée sur l'ensemble des données. Elle livre une «trace» qui traverse l'ensemble de la série. Une deuxième régression est calculée avec les points situés au-dessus de cette trace: cette régression livre le «plafond» du corridor. Une troisième régression calculée avec les points au-dessous de la trace donne le «plancher» du corridor. Le corridor défini par son plancher et son plafond est rectifié, transformé en un rectangle. Ce rectangle est obtenu par un centrage associé à une réduction de chaque largeur de cerne. Le centrage est obtenu par la soustraction de la valeur de la trace à l'année considérée et la réduction par la division du résultat centré par l'écart entre plancher et plafond de la même année. L'avantage de cette transformation sur d'autres, comme la correction par une moyenne mobile (Baillie et Pilcher) est qu'elle conserve une part d'ondulations mineures de la croissance qui, selon nous, sont susceptibles de caractériser plus précisément les différents environnements que la très haute fréquence.

les matrices sont assez «noires». Dans certaines d'entre elles, les plus basses corrélations ont été filtrées (évacuées), pour la lisibilité.

Le contraste entre zones très sombres et zones claires met en évidence des groupes de séries fortement apparentées – des familles homogènes «noires» – par opposition à des plages d'individus moins bien corrélés «blanches». Le rangement de la matrice joue un rôle efficace dans les capacités d'interprétation de la matrice. Dans le processus de datation, le rangement le plus efficace consiste à mettre en évidence la réplique des meilleures corrélations. Un algorithme du logiciel utilisé, inspiré des principes de Pareto ou loi des 20-80³, fait «monter» en haut à gauche les couples de séries qui présentent les plus fortes corrélations en ignorant les plus faibles. Ce type de rangement permet de distinguer rapidement des autres, les individus bien corrélés, bien «datés». Mais, sous l'angle de ce travail, ce type de rangement a un inconvénient. En effet, si la matrice contient plusieurs «familles» bien distinctes mais fortement homogènes, le rangement «de Pareto» a tendance à mélanger, en haut et à gauche, les meilleurs couples des différentes familles. Cette situation s'illustre par un certain désordre dans l'alternance des passages sombres et clairs le long de la diagonale. Il est alors pertinent de tenter d'améliorer le rangement de la matrice dans la perspective qui nous intéresse ici.

Parmi les différentes façons de classer une matrice symétrique de corrélations, la «classification hiérarchique ascendante»⁴ tente de hiérarchiser les corrélations par «familles». Noter : dans le langage de ce type de calcul une bonne «proximité» ou une faible «distance» équivaut à une bonne corrélation. Ces termes seront indifféremment utilisés dans ce qui suit. La classification hiérarchique se fait par étapes. Elle part des corrélations les plus élevées, pour finir par les plus basses. Au départ, il n'y a pas de familles, il n'y a que des éléments «orphelins», et petit à petit les «orphelins» seront intégrés aux familles en cours de constitution. À la différence du calcul précédent, à chaque pas de sa progression, la classification hiérarchique se préoccupe de la «proximité» de l'élément en cours de traitement avec ceux déjà associés dans des familles. S'il est plus proche d'une famille que d'un autre orphelin, cet élément est ajouté à la liste de la famille. Dans le cas contraire, le calcul crée une nouvelle famille avec les deux «orphelins» concernés. Et ainsi de suite, de proche en proche, les familles elles-mêmes «s'agglomèrent» en familles plus grandes (mais de moins en moins «homogènes») pour finir par une seule

hypothétique grande famille qui est l'ensemble de la population. Les étapes sont les suivantes (fig. 4) :

1. point de départ, il n'y a pas de «famille», il n'y a que des «orphelins».
2. les deux éléments les plus fortement corrélés sont saisis pour constituer la première famille, soit la première «branche» de l'arbre «hiérarchique» (qui mémorise la suite des opérations). Cette branche est aussi la plus courte de l'arbre car la longueur d'une «branche», la «distance», est le contraire de la corrélation (ici : distance = 1 – corrélation, ou, distance = 1 – probabilité).
3. La corrélation traitée est retirée du tableau original.
4. Le logiciel cherche alors la meilleure corrélation restante parmi les orphelins restants.

Il y a deux cas de figure :

- a) Ou bien cette corrélation concerne un autre orphelin : le logiciel crée une nouvelle famille avec les deux orphelins concernés. Une nouvelle branche indépendante de l'arbre est créée.
 - b) Ou bien cette corrélation concerne l'un des deux individus de la famille déjà trouvée. Dans ce cas la famille est augmentée d'un membre par «agglomération» de l'élément en cours de traitement. La famille compte alors 3 individus. Une nouvelle branche est ajoutée à la famille.
 5. La corrélation qui vient de travailler est effacée.
 6. Recommencer au point 4.
- Et ainsi de suite («boucler»).

Dans la «boucle» des opérations, à plusieurs reprises, il arrivera que la meilleure corrélation restante concerne les bords de deux familles déjà constituées, elles seront alors «agglomérées» en une super famille. Leurs arborescences sont alors liées par une branche et un «nœud» d'agglomération et les listes des deux familles sont juxtaposées.

À la fin tout le monde est «aggloméré», il n'y a plus qu'une famille et donc une seule liste. Cette liste donne l'ordre de la matrice nouvellement rangée. En parallèle, l'arbre hiérarchique qui mémorise la suite des opérations a été dessiné.

Il faut bien comprendre que les familles progressivement regroupées sont de moins en moins homogènes. Il reviendra à l'opérateur de décider des «nœuds» de partition pertinente. Le niveau de corrélation de ces nœuds peut varier d'une branche à l'autre car le calcul n'est pas dogmatique, il est descriptif, indicatif.

Dans ce papier, les arbres hiérarchiques se développent de gauche à droite, sauf le dernier (fig. 15) qui a été inversé pour des raisons graphiques.

3. Pareto 1909 et 1982.

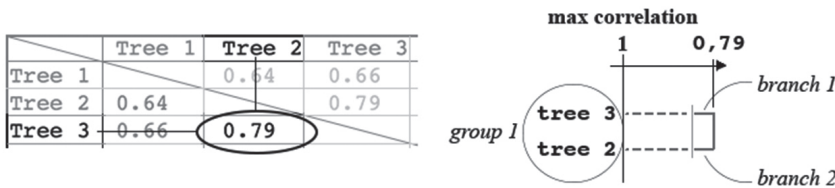
4. Jambu, Lebeaux 1978, Lerman 1981.

How to build of a typological hierarchical tree (a mathematical tree) Exemple with 5 oak trees of the Chaux forest, (Jura-39, France)

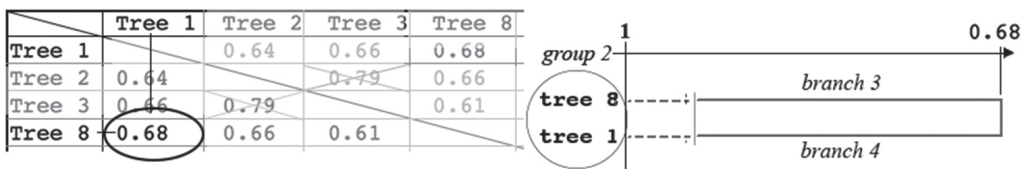
Phase 1 : do compute all inter-corrélations and to show them in a square matrix model

	Tree 1	Tree 2	Tree 3	Tree 8	Tree 9
Tree 1		0.64	0.66	0.68	0.31
Tree 2	0.64		0.79	0.66	0.44
Tree 3	0.66	0.79		0.61	0.48
Tree 8	0.68	0.66	0.61		0.47
Tree 9	0.31	0.44	0.48	0.47	

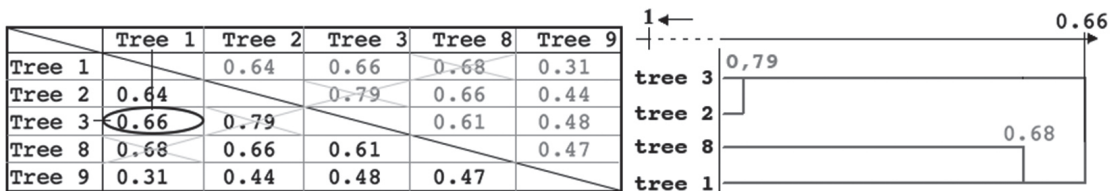
Phase 2 : do choose the best correlation and remove it from the board
then link the corelated elements as the first branches of the typological tree



Phase 3 : do choose the next best correlation, remove it from the board
then represent the correlation as new branches of the typological tree
In our case, the new branches 3 and 4 are not linked with the branches 1 and 2



Phase 4 : repeat the same operations. Now, the found correlation
between oak trees 1 and 3 links the mathematical branches which were previously built.



After the last phase : the mathematical tree is complete

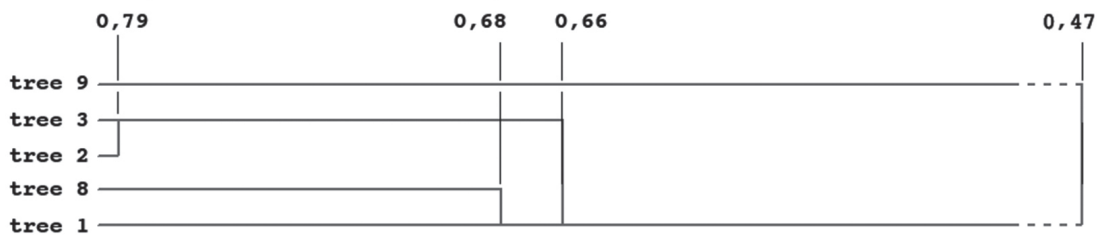


Fig. 4 – Principe du calcul d’un arbre hiérarchique ascendant ou arbre typologique. Un arbre ou une arborescence hiérarchique – souvent appelée arborescence typologique – est un graphique qui, à partir de critères quantifiés connus, tente de distinguer des groupes, des «familles» différents. Le principe consiste à calculer les «proximités» qui «rapprochent» les éléments les uns des autres, puis à les rassembler successivement, en commençant par les plus «proches» et en terminant par les plus «éloignés». Ici les «proximités» sont les corrélations ou les probabilités (de ressemblance) déduites du test t de Student. Il est impossible de construire un graphique qui montre toutes les proximités qui se développent dans un hyperespace à plus de 3 dimensions. Aussi, l’agglomération hiérarchique livre un point de vue sur cet espace par un arbre plan. Un arbre en mathématiques est un graphique dont deux points quelconques ne sont reliés que par un seul chemin. Ici, le chemin est une «branche» ou une suite de branches liées par des «nœuds» successifs. Les éléments les plus ressemblants sont reliés par les «branches» les plus courtes. À un niveau de ressemblance donnée (à une corrélation plancher par exemple) le calcul est capable de distinguer des familles différentes et indépendantes. Au cours du développement de l’arbre, les familles d’abord isolées seront finalement reliées à d’autre part des branches «longues». Il en sera de même des éléments les plus différents qui seront reliés à l’arbre par les «branches» les plus longues. L’objectif est d’évaluer la pertinence des familles de l’arbre au regard de la question posée, la proximité des arbres au sol, en «remontant» l’arbre des branches les plus courtes («petites» familles les plus cohérentes), aux branches les plus longues («grandes» familles les moins cohérentes).

La matrice, alors rangée dans l'ordre de la liste de l'arbre hiérarchique fait le bilan de l'algorithme et résume les résultats. Le «noyau» des familles y est en général bien représenté par des «pâtés» noirs (hautes corrélations) et les nœuds d'agglomération de l'arbre par des passages clairs (faibles corrélations) qui sont des chutes de corrélations consécutives.

Une matrice toute «noire» signale qu'il n'y a qu'une famille. Une matrice «claire» signale qu'il n'y a pas vraiment de famille. Un noyau noir dans un ensemble plus clair signale une famille qui comporte des éléments typiques («noirs») et d'autre plus libres ou alors étrangers mais trop dispersés pour constituer une autre famille (un deuxième noyau «noir»).

Le chercheur dispose alors de deux schémas de lecture de sa population d'arbres (vrais) : la matrice, synthèse des résultats, et l'arbre hiérarchique qui détaille la construction progressive des «familles» ou sous-groupes de la population analysée. L'analyse de l'arbre hiérarchique est une source de multiples observations intéressantes quand on analyse dans le détail une population donnée. Ici, il s'agit de donner une direction générale, les exemples qui suivent seront seulement utilisés dans leur principe général. Mais il y aurait beaucoup à dire de certaines figures...

1. ANALYSE D'UN ESPACE LOCAL : LA FORÊT DE CHAUX

La Forêt domaniale de Chaux est un vaste espace dans lequel il a été possible de faire des observations grâce à la bienveillance de l'administration des Eaux et Forêts (à l'époque). Des prélèvements groupés entre 10 et 30 arbres par placette ont été complétés par quelques échantillons qui ont pu être pris dans les environnements de travail. Une première phase a été faite sous la direction de l'auteur dans les années 1980 et une seconde par Olivier Girardclos (CNRS, Besançon) dans les années 2010. Au total, 16 stations sont documentées. Nous observerons la période 1910-1980.

1.1. Analyse des 16 moyennes de parcelles

Commençons par analyser les moyennes de parcelles (1 moyenne par parcelle). En Forêt de Chaux, le signal est assez homogène malgré les cycles forestiers qui ne frappent pas les parcelles en même temps, (fig. 5). On note une

remarquable chute de croissance en 1920 et 1921 (sécheresse de 1921), une autre vingt ans après et une oscillation en dents de scie des années 1956 – 1960, peut-être due au traumatisme dû à la vague de froid de février 1956 qui a altéré la périphérie des troncs. A priori, le terrain d'étude n'est pas favorable à une forte puissance de discernement géographique par le signal dendrochronologique, ce que confirme la matrice de corrélation des «moyennes» de station (fig. 6). Mais, dans la perspective d'une transposition historique, cette étendue correspond approximativement à la zone d'approvisionnement ordinaire d'un bourgeois aisé de la Renaissance rénovant sa propriété. Nous sommes donc dans un cas transposable dans une étude dendro-archéologique normale, avec ses contraintes.

L'arbre hiérarchique permet de ranger les séries par typologies voisines (fig. 7) et de ranger la matrice (fig. 8). L'arbre hiérarchique rapproche immédiatement les parcelles 96 et 97 (corrélation supérieure à 0,8) ainsi que les parcelles contiguës 212 et 213 (corrélation supérieure à 0,57). La parcelle 95, voisine des 96 et 97, a «attendu» d'être liée à celle dite «réserve» (éloignée) avant de se rallier à ses voisines (corrélations égales ou supérieures à 0,36). Ces coefficients de corrélation au-dessus de 0,35, sont très significatifs pour les 70+1 ans corrélés (1910 – 1980). Ce qui nous intéresse alors ce sont les écarts relatifs entre les nœuds d'agglomération des familles pour décider de la coupure entre deux familles pertinentes. Coupons au niveau de la corrélation 0,50 (fig. 8, ligne rouge) ce qui isole une première super famille de 5 éléments fortement corrélés (P96 – P97, puis P95 et P72), En pratiquant une autre coupure au niveau de corrélation 0,40, on étend la famille aux parcelles P212 – P213. Sont ainsi associées deux familles fortement «cousines» car les corrélations de coupure sont très élevées. La figure 9 suggère quelques détails du signal qui identifient les familles. Cet ensemble se détache très clairement dans la matrice de corrélation par un carré quasiment noir (groupe de corrélations maximales). Nous sommes en droit de supposer que les parcelles devraient être des plus rapprochées. La figure 10 montre qu'il n'en est rien, seules les familles élémentaires P96 – P97 et P212 – P213 sont réellement voisines au sol. Les coupures ne fonctionnent donc pas aussi bien que nous l'aurions aimé, à cause sans doute de la typologie des sols dont la spatialisation ne suit pas une géométrie régulière. Il faudrait étudier ce point précis, mais il n'y a pas la place ici. Pourtant l'arbre hiérarchique a bien distingué les stations nord-orientales des stations sud-occidentales.

Chaux Forest (Jura, France), Oak tree growth, averages of 16 places
 Period 1910 - 1980 AD (CE), corridor index.

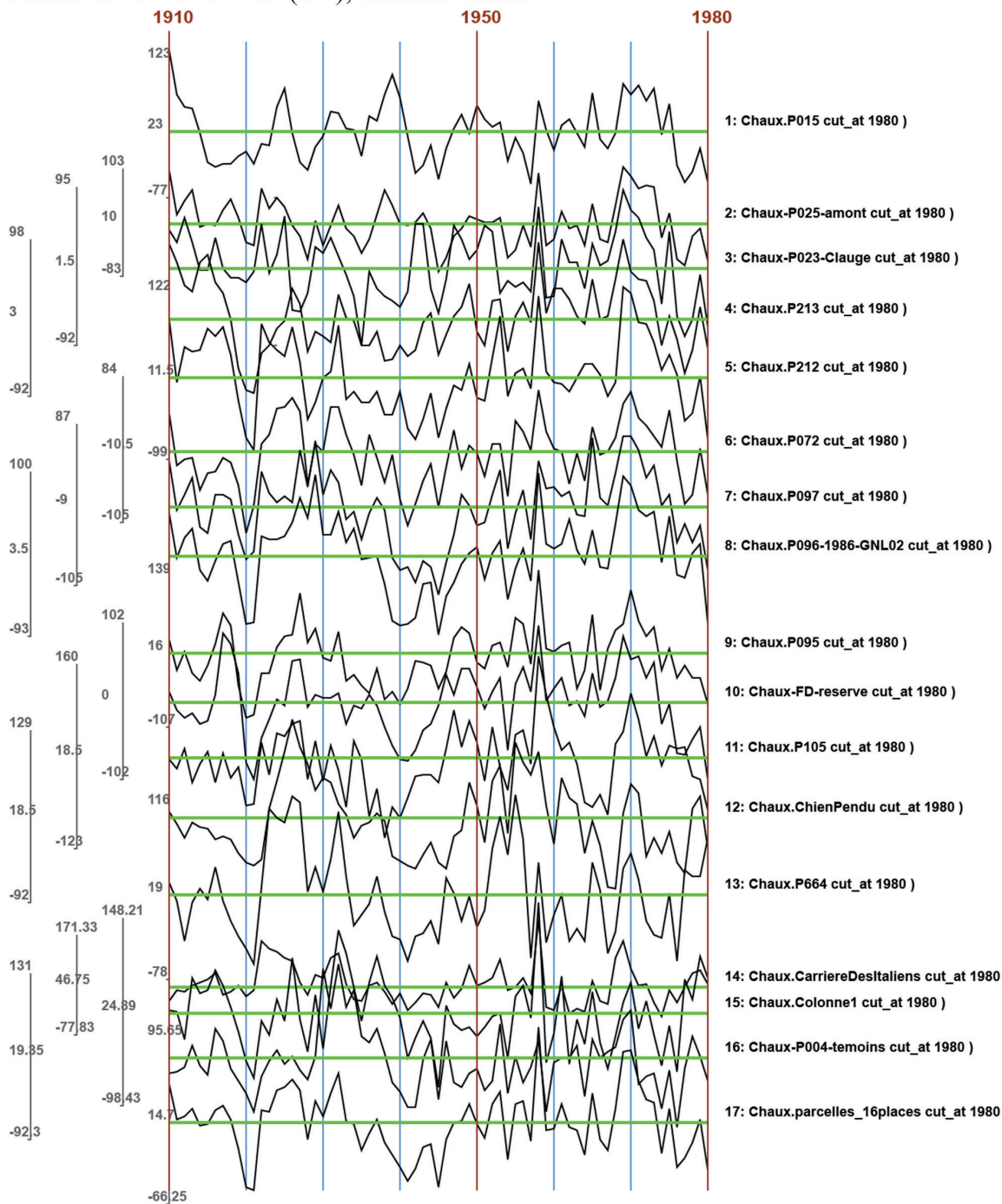


Fig. 5 – Graphique de synchronisation des moyennes dendrochronologiques des stations ou « parcelles » de la Forêt de Chaux dans le département du Jura (France) et réparties sur un espace d’une vingtaine de kilomètres. Période 1910-1980, essence analysée: le chêne. Noter immédiatement la forte convergence des mouvements de croissance de détail, quasi interannuels, entre les diverses parcelles. Ces mouvements sont la marque du climat. Les mouvements d’ensemble, à moyenne fréquence, se développant sur 10 à 30 ans, montrent plutôt la vie de ces parcelles: reprises de croissance après les coupes forestières puis chutes de croissance dans la reprise de la compétition naturelle pendant que les parcelles se reboisent. La calibration par le corridor réduit, mais conserve, ces mouvements de moyenne fréquence. Le corridor conserve donc la trace de la personnalité de chaque espace. Cette trace est le facteur qui permettra aux calculs de distinguer les différents lieux.

Chaux Forest, correlation matrix of the 16 local averages

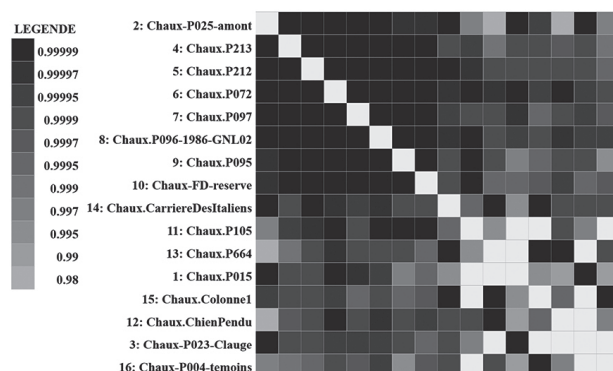


Fig. 6 – Matrice des corrélations croisées des moyennes de station de la Forêt de Chaux entre 1910 et 1980. En haut, à gauche, les neuf premières stations sont très fortement corrélées. Il est intuitif de supposer qu’il sera difficile de les différencier. Par contre, les sept parcelles suivantes signalent, à la fois, leur part d’originalité par la légère – mais perceptible – baisse de corrélation avec l’ensemble précédent (zone grisée en bas à gauche) et aussi leur indépendance respective par l’éclaircissement des inter-corrélations qui les concernent au voisinage de la diagonale (en bas, à droite) de la matrice. À ce stade la matrice n’a pas encore été rangée par l’arbre hiérarchique.

Chaux Forest, arrangement of the growth averages by a Hierarchical (typological) Tree

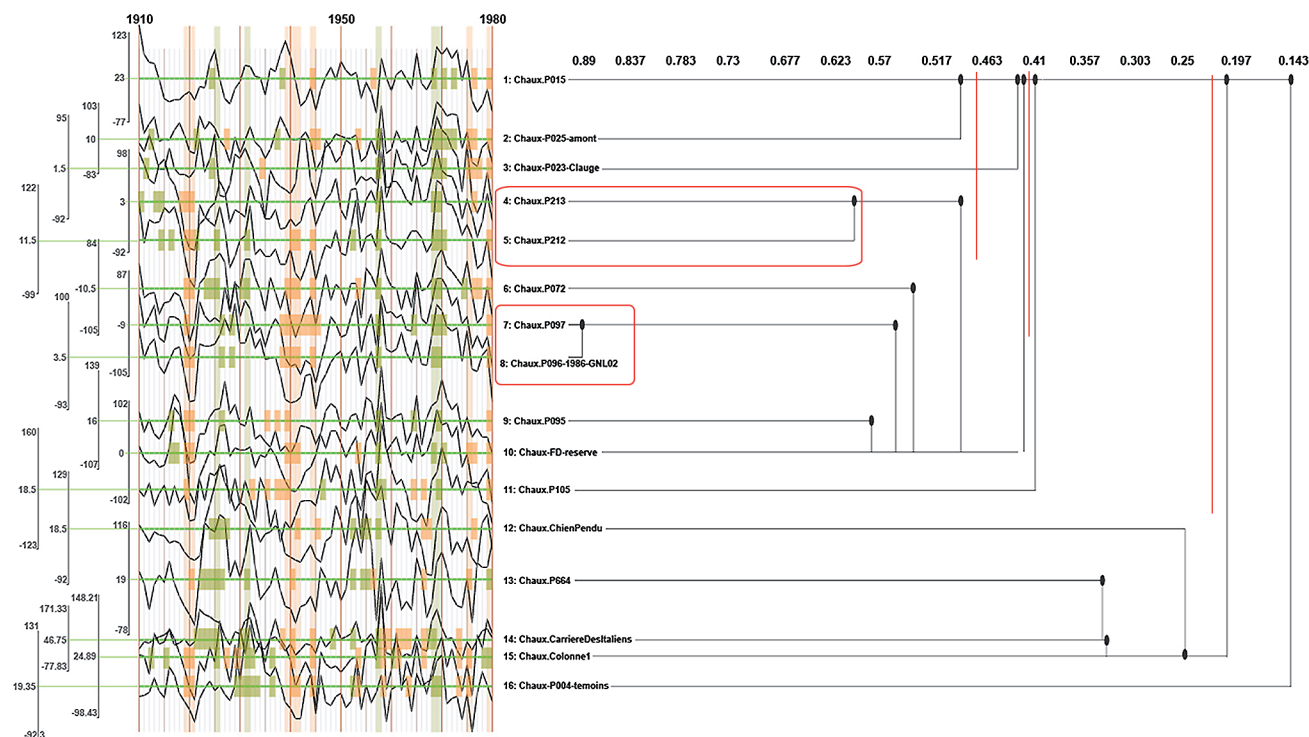


Fig. 7 – Arbre hiérarchique des corrélations des 16 stations de la Forêt de Chaux. L’arbre affiche 1 famille de stations très fortement corrélées : les parcelles 96 et 97 (probabilité d’identité voisine de 1, soit 0.99…), puis, le calcul suggère de rassembler aussi les parcelles 212 et 213 (probabilité voisine de 0.62) et, dans le voisinage, les parcelles 95 et « FD-réserve » (probabilité légèrement supérieure à 0.57). Ensuite, les probabilités significatives tombent rapidement, le lien possible ne s’impose plus.

Chaux Forest, averages of 16 places, years 1910 - 1980, typological identification on a matrix

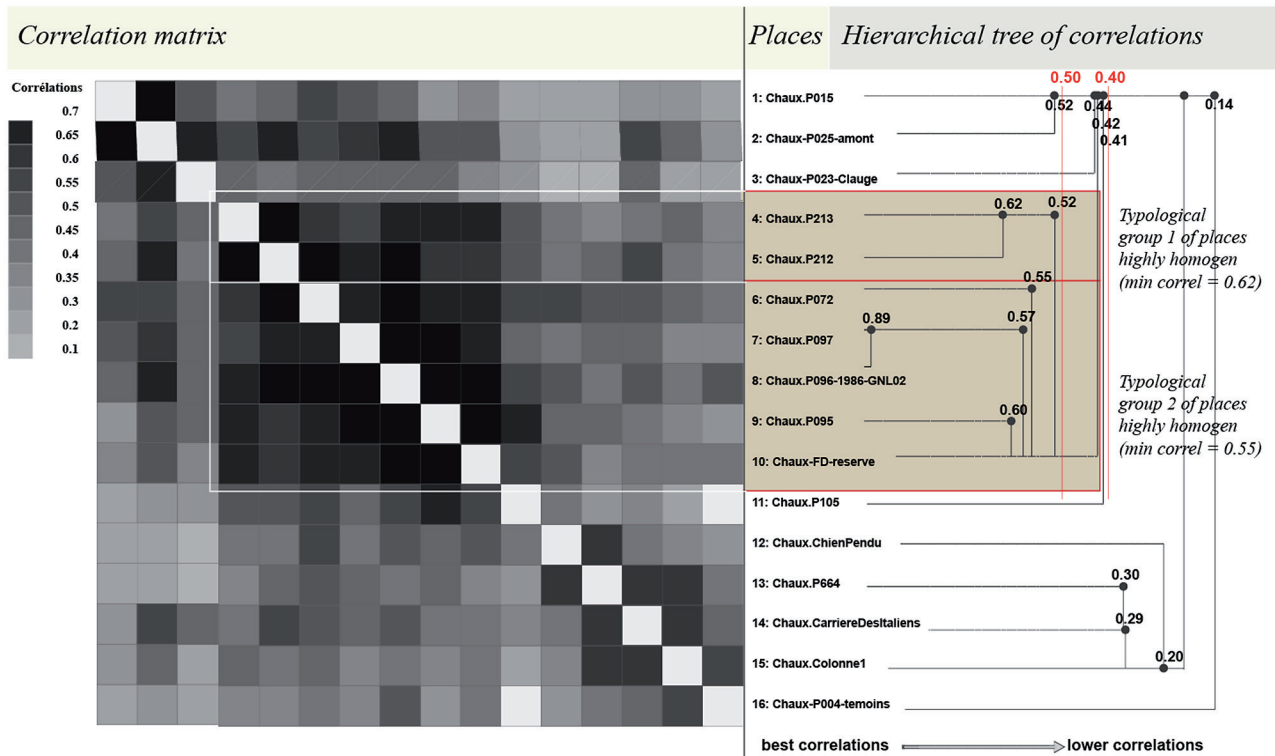


Fig. 8 – Rangement de la matrice des corrélations des moyennes de station par l'arbre et projection de celui-ci en marge droite de la matrice (Les branches les plus longues ont été compressées pour les besoins de la mise en page). La matrice rangée par l'arbre typologique permet d'en mieux interpréter les divisions. On y distingue le rassemblement en un pavé homogène de 7 des 9 stations déjà groupées en figure 6 (P212, P213, P97, P96, P95, P72 et FD-réserve). Par ailleurs, les stations hors de ce groupe pourraient être classées en deux sous-groupes bien distincts: un groupe du «haut», qui rassemblerait les parcelles 15, 23 et 25, et un groupe du «bas». En fait, nous verrons plus loin que l'arbre, indépendamment de la distance entre les nœuds a séparé les stations du nord-est de la forêt (zone plus drainée) de celles du sud-ouest (zone plus humide).

Chaux Forest, best typological groups, years 1910 - 1980, typological tokens

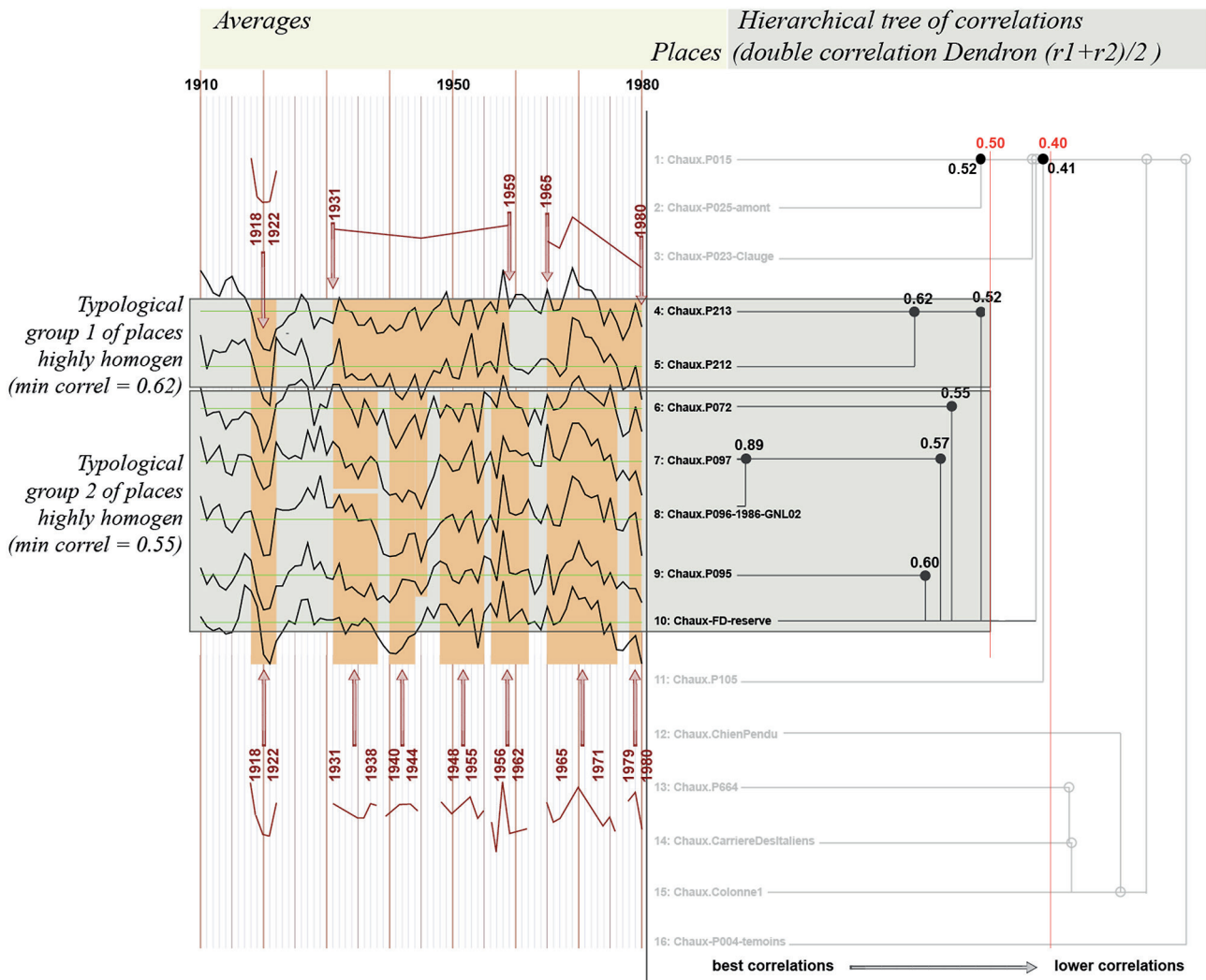


Fig. 9 – Analyse du groupe principal de stations de la Forêt de Chaux reconstitué par le calcul. Les croissances sont fortement parallèles. Une succession de marques du signal (signatures ou années significativement marquées) caractérise bien cet ensemble. Notamment les chutes de croissance des années 1918, 1922, 1954, 1976 et les reprises (météorologiques?) 1953, 1958 et 1969. Par contre la chute de croissance des années 1932 – 1945, sans doute attribuable à la concurrence liée à la reforestation, bien visible dans le sous-groupe principal n'a pas affecté les parcelles 212 et 213.

Chaux Forest, averages of 16 places, years 1910 - 1980, typological identification related to the locations

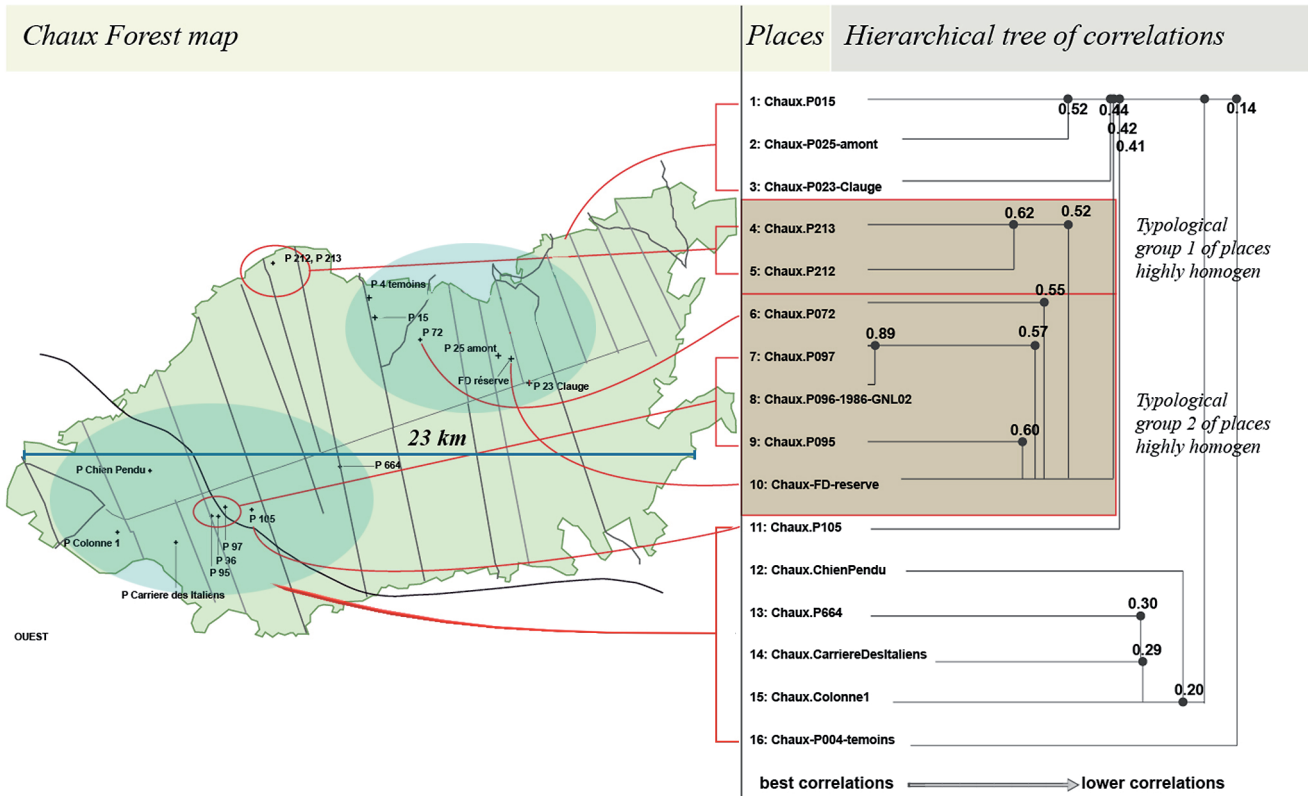


Fig. 10 – Projection sur le terrain des groupes proposés par l’arbre typologique et sa matrice de corrélations. En gros, l’arbre est parvenu à distinguer, les stations nord-orientales (haut de l’arbre) des stations sud-occidentales (bas de l’arbre) et plus en détail, il a rassemblé très clairement quatre des cinq stations contigües sur le terrain 212-213 et 96-97. Mais il a «éloigné» la station 95 de ses voisines (sur le terrain) 96 et 97, car il lui a trouvé une meilleure affinité avec la station éloignée FD-réserve. Affinité dont les causes sont sans doute pédologiques. Au final, il apparaît que le calcul, appliqué à des moyennes, a bien une prédisposition à discriminer des lieux, mais que des confusions, dues à des conditions écologiques semblables mais dispersées à cette échelle, ne sont pas exclues.

1.2. Analyse des 129 individus non regroupés

Appliquons les mêmes calculs aux individus après en avoir exclu quelques-uns qui «écrasent» l’arbre hiérarchique et sont à discuter dans un autre cadre. La matrice reclassée par l’arbre hiérarchique obtenu (fig. 11) n’améliore pas le résultat précédent. Les arbres des parcelles P95, P96, P97 sont dispersés dans l’arbre hiérarchique et donc dispersés dans la matrice, mais, inversement, ceux de trois autres parcelles (P664, «Colonne 1» et «Chien Pendu») sont tous rassemblés dans «leur» famille. Définissons une «performance» du rangement des arbres obtenus : la performance est la proportion d’éléments consécutifs qui représentent au moins 80% des arbres d’une parcelle rapportée au nombre total d’arbres calculés. La «performance» du calcul est donc ici de 37%, ce qui n’est pas suffisant. Réfléchissons.

1.3. Analyse d’un échantillon de la population

Il se pose la question de la représentation de chaque station forestière dans le calcul, certaines d’entre elles ayant fourni 20 arbres ou plus et d’autres n’en ayant fourni que deux ou trois. Cette différence donne un «poids» différent aux différentes placettes dans le calcul et d’une certaine façon, fausse l’égalité des chances dans la construction des nœuds d’agglomération. Il serait donc préférable de travailler sur le même nombre d’arbres pour chaque station. La station la moins «peuplée» impose alors ce nombre. Nous en prendrons cinq dans chacune des 11 stations suffisamment fournies. Censés représenter au mieux la station, ils seront choisis au plus près de la moyenne de station, soit, en fait, dans la zone la plus sombre de la matrice de corrélation de chaque station. Nous conserverons également les éléments des autres stations, à condition qu’il y en ait au moins deux. La figure 12 montre la position des séries choisies dans la matrice de chaque station. 67 séries sont ainsi rassemblées dans un nouvel ensemble à calculer.

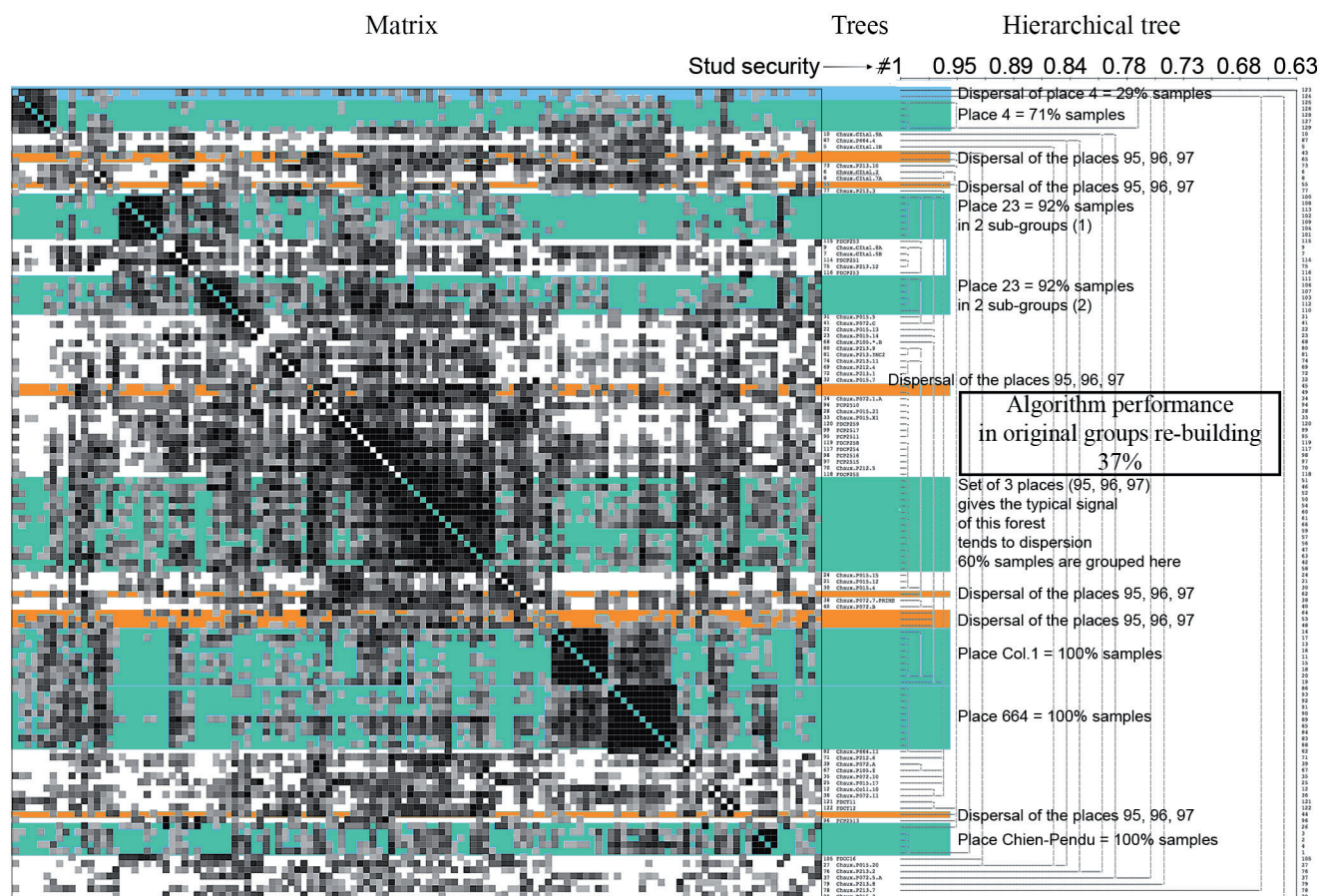


Fig. 11 – Analyse de tous les individus de la Forêt de Chaux dans la période 1910 – 1980, arbre hiérarchique et sa matrice de corrélations. Le calcul des individus est plus explicite que le calcul sur les moyennes. Les plages sombres de fortes corrélations et distinctes le long de la diagonale suggèrent un plus grand nombre de groupes potentiels. Peut-être six ou sept. Mais nous savons qu'il y a 16 stations différentes, le calcul aurait donc, intuitivement, un potentiel de discernement de l'ordre d'un tiers ($6/16$ ou $7/16$). Nous proposons de calculer plus précisément la «performance» de ce type de calcul (voir dans le texte). La performance est un nombre qui évalue le taux d'individus correctement rassemblés par l'arbre typologique. La performance est ici de 37 %, soit à peu près ce qui était pressenti intuitivement. Le brouillard reste donc un peu trop épais à notre avis, tentons de l'éclaircir.

On y ajoutera – pour voir – cinq bois d'un site voisin, situé à 15 km au sud-ouest, sur la commune d'Ouessières (Jura). Soit, au total, 72 séries à calculer.

Le résultat, figure 13, est alors bien plus intéressant que les précédents. Les séries de 11 placettes sont rassemblées à au moins 80 % et la performance du calcul, établie comme précédemment, est de 71 %. L'amélioration était attendue, car la concentration de l'échantillonnage autour de la moyenne des stations améliore l'homogénéité des groupes retenus et diminue leur porosité, leur aptitude à se laisser pénétrer par des éléments étrangers. Il resterait à étudier de plus près ce qui se passe au niveau de la pertinence des calculs quand ils passent d'une population entière à un échantillonnage de cette population comme nous venons de le faire. Mais pour étudier des populations plus importantes, il faudra bien échantillonner. C'est ce que nous allons faire avec une population d'arbres qui couvrent toute la France.

2. ANALYSE D'UNE VASTE RÉGION

La figure 2 montre la répartition des points à notre disposition en France tempérée, Belgique du Sud et Suisse occidentale. Matière qui nous est accessible grâce à l'amabilité de nombreux collègues cités en tête de cet article. Soit 32 stations réparties sur presque 800 km, du Causse de Martel à la région de Liège, de la Charente au Canton de Neuchâtel et de l'Isère à la Normandie (555 arbres). Nous avons déjà mobilisé cette base de données collective pour un papier traitant de la collaboration potentielle du même matériau à la dendroclimatologie⁵. Il s'agit uniquement de chênes, sessiles, pédonculés ou hybridés entre eux. La représentativité au sol de ce set de données est déséquilibrée en faveur de l'Est de la France, région bien mieux échantillonnée que le reste du pays. Il faut faire avec.

5. Lambert *et al.* 2010.

Select the 5 best correlated oak trees of each place = select best samples of the typical signal of each place

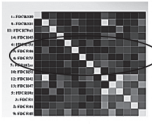
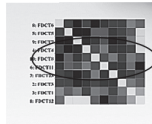
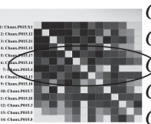
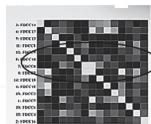
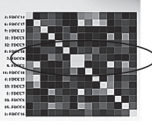
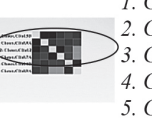
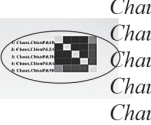
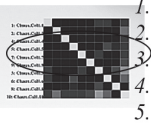

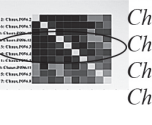
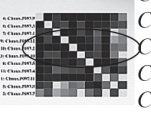
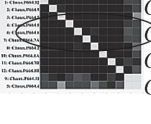
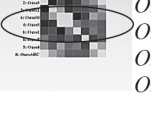
<p>Chaux FD réserve</p>  <ol style="list-style-type: none"> 1. FDCR46 2. FDCR51 3. FDCR73 4. FDCR75 	<p>Chaux P 4 témoins</p>  <ol style="list-style-type: none"> 1. FDCT11 2. FDCT13 3. FDCT4 4. FDCT7 5. FDCT8 	<p>Chaux P 15</p>  <p>Chaux.P015.13 Chaux.P015.14 Chaux.P015.17 Chaux.P015.18 Chaux.P015.4</p>	<p>Chaux P 23 «Clauge»</p>  <ol style="list-style-type: none"> 1. FDCC1 2. FDCC13 3. FDCC18 4. FDCC8 5. FDCC9
<p>Chaux P 25 «amont»</p>  <ol style="list-style-type: none"> 1. FCP2511 2. FCP2515 3. FCP2517 4. FDCP254 5. FDCP258 	<p>Chaux «Carrière des Italiens»</p>  <ol style="list-style-type: none"> 1. Chaux.CItal.1B 2. Chaux.CItal.2 3. Chaux.CItal.5B 4. Chaux.CItal.7A 5. Chaux.CItal.8A 	<p>Chaux «Chien Pendu»</p>  <p>Chaux.ChienPd.1B Chaux.ChienPd.2A Chaux.ChienPd.3B Chaux.ChienPd.8A Chaux.ChienPd.9B</p>	<p>Chaux «Colonne 1»</p>  <ol style="list-style-type: none"> 1. Chaux.Coll.3 2. Chaux.Coll.4 3. Chaux.Coll.5 4. Chaux.Coll.7 5. Chaux.Coll.9
<p>Chaux P 72</p>  <p>Chaux.P072.1.A Chaux.P072.5.A Chaux.P072.A Chaux.P072.B Chaux.P072.C</p>	<p>Chaux P 96</p>  <p>Chaux.P096.1 Chaux.P096.10 Chaux.P096.12 Chaux.P096.3 Chaux.P096.4</p>	<p>Chaux P 97</p>  <p>Chaux.P097.1 Chaux.P097.10 Chaux.P097.12 Chaux.P097.2 Chaux.P097.5</p>	<p>Chaux P 105</p> <p>(too few trees : no matrix)</p> <p>Chaux.P105.*.B Chaux.P105.S</p>
<p>Chaux P 212</p> <p>(too few trees : no matrix)</p> <p>Chaux.P212.4 Chaux.P212.5</p>	<p>Chaux P 213</p> <p>(too few trees : no matrix)</p> <p>Chaux.P213.1 Chaux.P213.10 Chaux.P213.11</p>	<p>Chaux P 664</p>  <p>Chaux.P664.5 Chaux.P664.6 Chaux.P664.7A Chaux.P664.8 Chaux.P664.9</p>	<p>Oussières 1982</p>  <p>Ouss1 Ouss10 Ouss12 Ouss3 Ouss5</p>

Fig. 12 – Préparation d'un échantillon «calibré» du même ensemble. On entend par «calibré», un échantillonnage régulier des parcelles : c'est-à-dire prendre à peu près le même nombre d'arbres par parcelle. Après analyse de chaque parcelle, les arbres seront choisis parmi les plus caractéristiques du signal moyen de la parcelle ou, plus simplement, dans la zone de hautes corrélations la plus «noire» de sa matrice de corrélation. Cinq arbres maximum par parcelle seront retenus, celles qui ont fourni moins de 5 individus resteront représentées par tous leurs arbres. 67 séries seront ainsi retenues pour le calcul suivant.

La Forêt de Chaux précédente y est représentée par trois moyennes de synthèse utilisées dans un travail précédent.

2.1. Analyse des moyennes de sites dans une vaste région

La figure 14 montre la matrice de corrélation des moyennes de site après rangement par l'arbre hiérarchique. La période calculée est toujours 1910-1980. Considérons l'aptitude de l'algorithme à distinguer les deux régions Est (en orangé sur la figure) et Ouest (sans couleur). Une différenciation des deux régions est effectivement esquissée. Les sites de l'Est marquent une préférence pour le haut de la matrice (11 sites groupés sur 15), ceux de l'Ouest pour le bas. Mais le bloc du haut ne donne pas 80% des sites de l'Est (11/15 = 73%), ce qui donne une performance de l'algorithme de zéro, au sens où elle a été définie précédemment. Il apparaît donc, une fois de plus, qu'au regard de la dendro-provenance, ce type de calcul ne doit pas manipuler

des moyennes archéologiques (moyennes globales de sites) en l'état. Il faudrait au moins les resserrer typologiquement et réduire la variabilité autour du signal moyen. Cette remarque se trouve confortée par le test qui suit.

2.2. Analyse d'un échantillonnage d'individus dans une vaste région.

Le nombre d'échantillons prélevés est très différent d'une station à l'autre, il passe d'une douzaine à une quarantaine. Les mêmes remarques que sur le plan de prélèvement de la Forêt de Chaux, s'appliquent ici. Il est préférable de travailler sur des nombres d'arbres identiques dans toutes les stations. Nous reprendrons le même type d'échantillonnage que précédemment, soit 5 arbres par forêt, également choisis parmi les plus corrélés à leur moyenne de site, ou à leur moyenne de «famille», si le site comporte plusieurs groupes (ce qui est le cas de la Forêt de Chaux). Sélection qui donne 155 arbres à calculer.

Chaux Forest : Computing with a population of 5 significant trees taken in each place

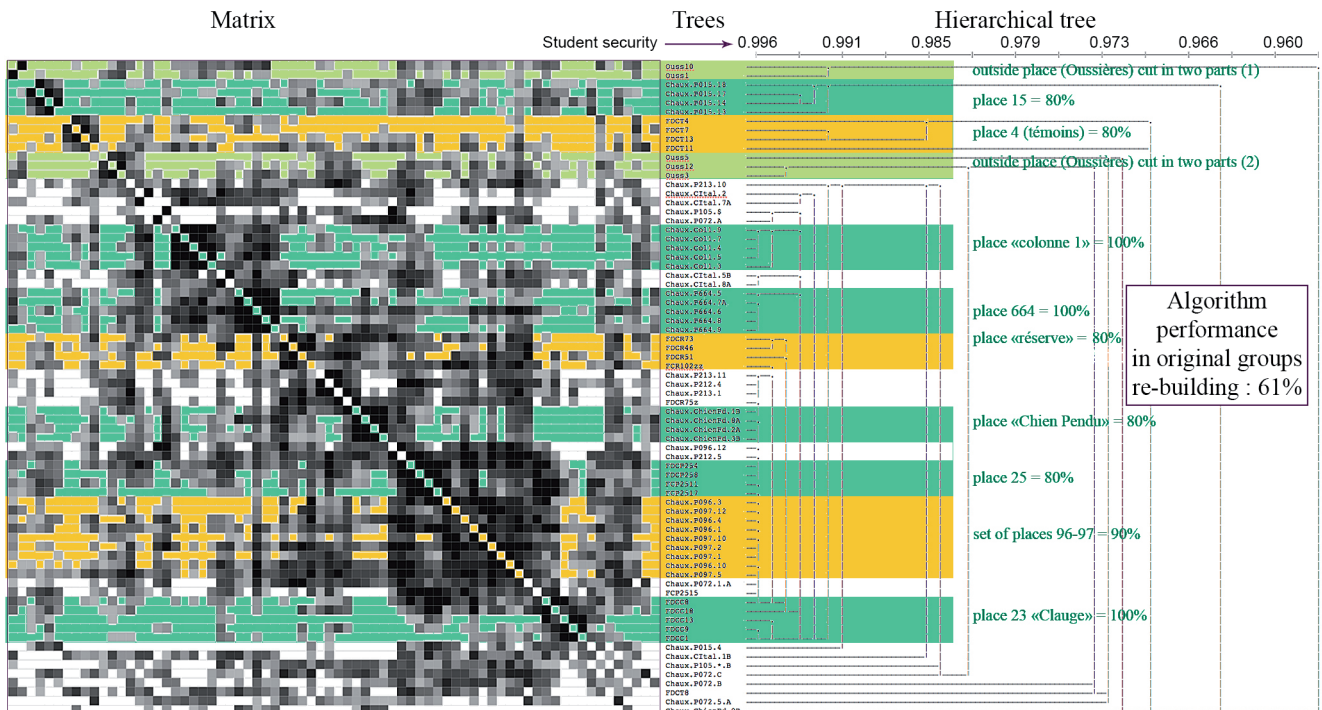


Fig. 13 – Analyse de l'échantillon de population composé des 67 arbres triés selon les matrices de parcelles. Un nouvel arbre hiérarchique est calculé sur ce nouvel ensemble ainsi que sa matrice de corrélations. On obtient un résultat semblable en apparence à celui de la figure 11. Cependant, le groupement des individus par station s'est amélioré. C'est assez logique car nous avons en quelque sorte «forcé» le signal de chacune d'elles en ne conservant que les séries les plus typiques du signal supposé standard pour chacune d'elles. Le score est meilleur que celui de la figure 11, en effet la «performance» qui évalue le taux de rassemblements corrects passe à 61 %. Soit, en raccourci, le calcul serait apte à rassembler plus de la moitié des bois analysés.

Try to identify and separate eastern and western forests using the averages of the places

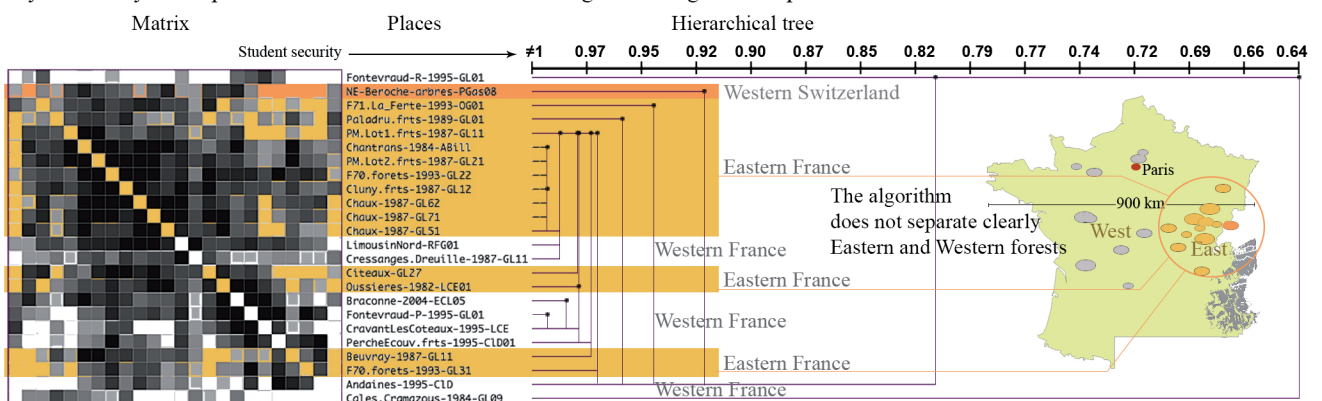


Fig. 14 – Analyse des moyennes de 24 stations réparties sur une vaste zone couvrant la France tempérée. Ici, une question est posée à l'arbre hiérarchique et sa matrice, à savoir s'ils sont capables de distinguer les stations de l'Est, continentales, des stations de l'Ouest, plus océaniques. La réponse du calcul est mitigée. Il a en effet rangé en haut de la matrice (et en haut de l'arbre typologique) une majorité de stations de l'Est, soit 11 stations. Mais quatre stations de l'Est – soit près de 27% – (Citeaux (21), Beuvray (71), Oussières (39) et F70-forêts de Haute-Saône) se trouvent mêlées aux stations de l'Ouest. La discrimination est/ouest n'est donc positive qu'à 73%. Mais nous avons déjà vu, figures 6 et 7, que les moyennes ne semblent pas être de bons vecteurs pour porter la question de la provenance : il est préférable de travailler sur les individus.

Sample of 155 trees extracted from a set of 32 forests spread in northern France, southern Belgium and western Switzerland

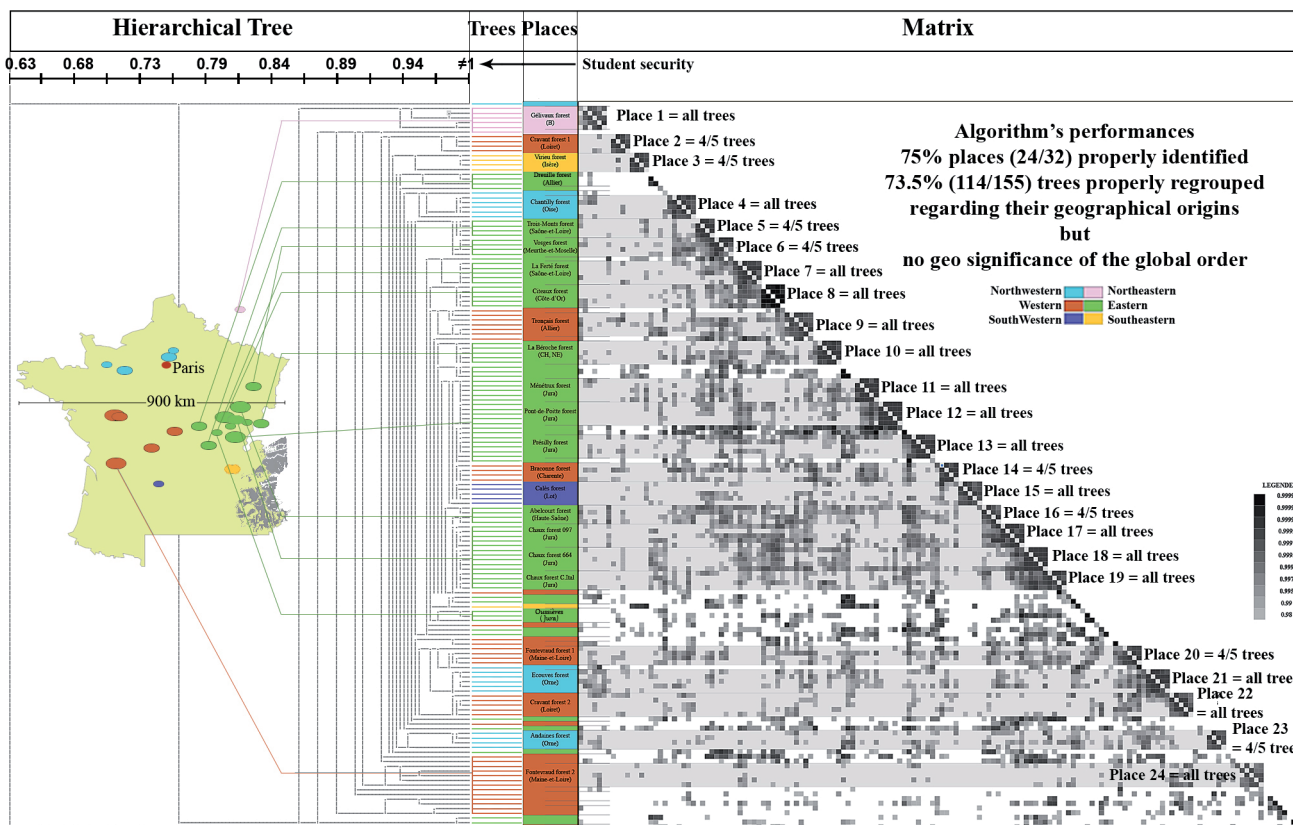


Fig. 15 – Analyse d’un échantillon de la population globale du chêne tempéré entre France, Belgique et Suisse occidentale. Basé sur le dernier calcul de la Forêt de Chaux, l’échantillonnage a consisté à prendre 5 arbres par station ou placette censés représenter au mieux le signal de station. La méthode de sélection a été la même que celle décrite figure 12. L’arbre hiérarchique et sa matrice de corrélations ont été appliqués aux 155 arbres sélectionnés. Pour simplifier la figure, seule la demi-matrice est représentée et l’arbre placé à gauche se lit en sens inverse des précédents, de droite à gauche. Les sites de l’Est, les plus nombreux, ont tendance à occuper le milieu de la figure et à rejeter sur les bords, en haut et en bas, des sites dont le signal est différent. Le résultat est intéressant car 24 places sur les 32 (75%) ont été identifiées par au moins à 80% de leurs individus. La performance de regroupement des individus atteint 73,5%. On considérera ces derniers pourcentages comme une base d’appréciation provisoire – mais bien factuelle – de la capacité de discrimination géographique de la méthode.

Les mêmes calculs sont engagés : matrice de corrélation de base, arbre hiérarchique et réarrangement de la matrice par l’arbre. La figure 15 montre un résultat intéressant : 75 % des places ont été correctement recomposées (en regroupant au minimum 80 % des arbres d’une station) ce qui a pour conséquence que 71 % (114 / 155) des arbres de la population ont été correctement rapprochés par l’algorithme. On peut suivre le long de la diagonale la succession des points de corrélations intra-station dont 15 d’entre eux rassemblent en un seul bloc les cinq arbres retenus pour représenter une station. Il semble que la distance – unique paramètre environnemental dont il est tenu compte ici – améliore la capacité discriminante du calcul et que celui-ci peut être utilisé comme discriminateur de lieux dans le cadre d’une sécurité supérieure à 70%.

Par contre l’ordre général de rangement n’a pas de sens devant la question posée.

CONCLUSION

Le processus peut se résumer ainsi :

1. calcul de l’indice écrêté du corridor de chaque série ;
2. utilisation du Coefficient de Corrélation ou de la probabilité de Student comme «distance» typologique ;
3. utilisation des individus et non pas des moyennes ;
4. constitution de populations homogènes ;
5. évacuation des individus marginaux ;
6. échantillonnage des groupes repérés à la datation ;
7. calcul de la matrice d’inter-corrélation de l’échantillon global ;
8. calcul de l’arbre hiérarchique à partir des corrélations ou des probabilités de Student ;
9. reclassement de la matrice par la classification hiérarchique ascendante ;
10. repérage des nœuds de corrélation de l’arbre typologique potentiellement pertinents et/ou délimita-

tion des taches de forte corrélation sur la diagonale de la matrice reclassée susceptibles de circonscrire des «familles» géographiques;

11. hypothèses de provenances en mobilisant l'environnement scientifique de l'étude.

Les calculs standards de datation peuvent être mobilisés pour produire des graphiques utilisables comme bases d'hypothèses de discrimination de lieux d'origine des arbres analysés en dendrochronologie. L'exercice développé ici, avec ses contraintes, débouche sur une pertinence potentielle de l'ordre de 75%. Résultat qui suggère que ce type de calcul doit être pris en considération dans une réflexion sur la dendro-provenance, mais il laisse une place légitime et assez large aux interprétations du chercheur.

Note : tous les calculs ont été faits avec le logiciel expérimental Dendron-IV⁶.

Bibliographie

- Jambu, Lebeaux 1978 :** M. Jambu, M.-O. Lebeaux M.-O., *Classification automatique pour l'analyse des données*, Dunod, Paris, 1978, 2 vols.
- Lambert 2006 :** G. Lambert, *Dendrochronologie, histoire et archéologie, modélisation du temps ; le logiciel Dendron II et le projet Historic Oaks*, Habilitation à Diriger les Recherches (HDR), Université de Franche-Comté, Besançon, 2006, 2 vol.
- Lambert et al. 2008 :** G. Lambert, B. Saint-Jean-Vitus, J.-L. Taupin, Y. Trenard, La datation dendrochronologique est-elle une science ? Une question de vocabulaire. Exemples du Tournugeois (71) et du Petit Louvre de La Pacaudière (42). *Structures bois dans le patrimoine bâti. ICOMOS, Actes des journées techniques internationales Bois, Metz, 29, 30 et 31 mai 2008*, 22-27.
- Lambert et al. 2010 :** G. Lambert, V. Bernard, J.-L. Dupouey, P. Fraiture, P. Gassmann, O. Girardclos, F. Lebourgeois, Y. Ledigol, C. Perrault, W. Tegel, Dendrochronologie et Dendroclimatologie du chêne en France. Questions posées par le transfert de données de bois historiques vers la Dendroclimatologie, in L. Astrad, C. Miramont (dir), *Panorama de la Dendrochronologie en France*, Université de Savoie, collection Edytem, 11, 2010, 205-216.
- Lerman 1981 :** I.-C. Lerman, *Classification et analyse ordinale des données*. Dunod, Paris, 1981, 740 p.
- Pareto (1896) 1909 :** V. Pareto, *Manuel d'Economie Politique*, V. Giard et E. Brière éditeurs, Paris, 1909.
- Pareto (1894) 1982 :** V. Pareto, Premier cours d'économie politique appliquée, donné à l'Université de Lausanne in Pareto Vilfredo, *Œuvres complètes*, t. XXIV, Droz, Genève, Travaux de Sciences Sociales, (1894), 1982, 264 p.

6. Lambert 2006, Dendron II / Dendron IV.