

STATISTIQUE, BIOMÉTRIE, AGRONOMIE: APPROCHE HISTORIQUE⁽¹⁾

Pierre Dagnelie

Faculté des Sciences agronomiques
B-5030 Gembloux (Belgique)

pierre@dagnelie.be

RÉSUMÉ

Cette note retrace l'évolution de la statistique et de la biométrie, en relation avec l'agronomie, au cours des dix-neuvième et vingtième siècles. Elle met notamment en évidence les liens étroits qui ont existé entre ces disciplines durant la première moitié du vingtième siècle et l'influence prépondérante du développement de l'informatique au cours de la seconde moitié de ce siècle.

SUMMARY

This note sketches the evolution of statistics and biometry, in relation with agronomy, during the nineteenth and twentieth centuries. In particular, it stresses the close links between these disciplines during the first half of the twentieth century and the main influence of the development of computer science during the second half of this century.

1. INTRODUCTION

Depuis bien longtemps, la statistique et la biométrie sont des éléments essentiels de la recherche agronomique. La situation a sensiblement évolué au cours des dernières années, sans que diminue cependant l'importance de ces piliers de la recherche. Aussi nous a-t-il paru opportun de faire le point en la matière.

Dans un premier temps, je voudrais retracer l'évolution de la statistique et de la biométrie, en relation avec l'agronomie, du dix-neuvième siècle à nos jours. Ensuite, Elisabeth **de Turckheim** nous parlera de ce que peuvent être, à l'heure actuelle, la politique et les thèmes de recherche en biométrie, dans un institut de recherche agronomique. Ensuite encore, Jean **Tranchefort** nous présentera le rôle du biométricien consultant, les différentes facettes et les exigences de son métier, dans les instituts de recherche-développement. Enfin, je reprendrai la parole, en considérant alors l'avenir, plutôt que le passé.

En ce qui concerne ce premier exposé, j'envisagerai successivement **trois périodes**:

- le dix-neuvième siècle, marqué par le développement de la statistique et la naissance de la biométrie;
- la première moitié du vingtième siècle, et en particulier les relations entre l'agronomie, la statistique et la biométrie;
- et la seconde moitié du vingtième siècle, avec les faits majeurs que sont l'apparition de l'ordinateur et l'expansion de la micro-informatique.

Je tiens à préciser, en commençant, que le terme **agronomie** est considéré ici dans un sens très large, englobant tout ce qui a trait non seulement aux grandes cultures et à l'élevage,

(1) *Comptes Rendus de l'Académie d'Agriculture de France* **81** (8), 33-39, 1995.

mais aussi à l'horticulture, aux eaux et forêts, au génie rural, à l'économie et la sociologie rurales, aux industries agroalimentaires et aux biotechnologies, sans oublier les sciences de base sous-jacentes.

Mais avant de citer d'autres noms, je souhaite évoquer la mémoire de Jean **Dufrenoy** (1894-1972), ancien Membre de l'Académie d'Agriculture et ancien Président de la Société de Statistique de Paris, qui a souvent parlé de statistique et de biométrie, en ces lieux, de 1957 à 1972. Jean **Dufrenoy** a notamment publié, dans les *Comptes Rendus de l'Académie d'Agriculture*, diverses notes relatives à la théorie de l'information et à son utilisation en agronomie, aux problèmes de formation mathématique pour les recherches agronomiques et aux comparaisons multiples de moyennes. Il a aussi publié des travaux à caractère statistique ou biométrique, entre autres, dans les *Comptes Rendus de l'Académie des Sciences*, le *Journal de la Société de Statistique de Paris* et les *Cahiers des Ingénieurs agronomes* (41).

2. LE DIX-NEUVIÈME SIÈCLE: LE DÉVELOPPEMENT DE LA STATISTIQUE ET LA NAISSANCE DE LA BIOMÉTRIE

La statistique peut être définie comme étant l'ensemble des méthodes qui permettent de rassembler et d'analyser des données numériques – ou des informations codées sous forme numérique – à propos d'un ou plusieurs groupes d'individus (êtres, objets, etc.)⁽¹⁾.

La statistique moderne, telle que nous la connaissons aujourd'hui, est née durant le dix-neuvième siècle, de la fusion de différents courants antérieurs. Parmi ceux-ci, on peut citer l'arithmétique politique, visant à la connaissance des États, et le calcul des probabilités, initialement lié aux jeux de hasard (10, 11). Il me plaît de souligner à cet égard le rôle prépondérant d'Adolphe **Quetelet** (1796-1874), Belge que je qualifierais volontiers d'Européen, puisqu'il est né à Gand sous le régime français, qu'il a fait ses études et a commencé sa carrière à Gand et à Bruxelles sous le régime hollandais, et a ensuite très largement contribué au développement de la statistique – et de l'astronomie – tant en Belgique qu'à l'étranger (17, 26).

L'idée maîtresse de **Quetelet** était d'appliquer le calcul – ou la théorie – des probabilités "à toutes les sciences d'observation" (30). Tel est bien l'objet de la **biométrie**, généralement définie comme étant l'application des méthodes statistiques – ou des méthodes mathématiques, dont la statistique – aux sciences biologiques⁽²⁾.

La biométrie est née à la fin du dix-neuvième siècle, dans la ligne des travaux de Francis **Galton** (1822-1911) et Karl **Pearson** (1857-1936) notamment (19, 33). Le premier numéro de la revue *Biometrika* a d'ailleurs paru en 1901.

Dans le secteur agronomique, les mots *biométrie* et *statistique* sont souvent confondus ou, en tout cas, utilisés avec des significations très proches. L'INRA, par exemple, privilégie le mot *biométrie*. Personnellement, j'utilise plus volontiers le mot *statistique*, considérant qu'une part non négligeable des applications que nous rencontrons en agronomie appartiennent à d'autres domaines que la biologie. Tel est souvent le cas, entre autres, en génie rural et en économie rurale.

(1) Utilisé dans ce sens, le mot statistique devrait toujours figurer au singulier. Mais le même mot, employé tantôt au singulier, tantôt au pluriel, peut également désigner les données ou les informations elles-mêmes.

(2) Le mot biostatistique est parfois considéré comme synonyme de biométrie (21, 31).

3. LA PREMIÈRE MOITIÉ DU VINGTIÈME SIÈCLE: LES RELATIONS ENTRE L'AGRONOMIE, LA STATISTIQUE ET LA BIOMÉTRIE

La première moitié du vingtième siècle est marquée par un ensemble de relations très étroites entre l'agronomie, d'une part, la statistique et la biométrie, d'autre part. Nombre de statisticiens et de biométriciens illustres ont en effet oeuvré, au cours de cette période, dans l'enseignement et la recherche agronomiques.

À tout seigneur, tout honneur: Ronald Aylmer **Fisher** (1890-1962) est engagé en 1919 à la *Rothamsted Experimental Station*, au nord de Londres. Il y travaille jusqu'en 1933 et il pourrait bien être ainsi le premier mathématicien employé à temps plein dans un centre de recherche agronomique (6, 21).

Mais d'autres noms peuvent être cités. Je pense en particulier, dans l'ordre de leurs dates de naissance, à:

- William Sealy **Gosset** (1876-1937), plus connu sous son pseudonyme **Student**, qui travaille chez Guinness – nous dirions aujourd'hui "dans l'agroalimentaire" – à Dublin et Londres de 1899 à 1937 (5, 29);
- George Waddel **Snedecor** (1881-1974), à l'*Iowa State College* de 1913 à 1958 (9, 21);
- John **Wishart** (1898-1956), à Rothamsted de 1927 à 1931 et à la *School of Agriculture* de l'Université de Cambridge de 1931 à 1953 (3);
- Gertrude Mary **Cox** (1900-1978), à l'*Iowa State College*, au *North Carolina State College* et à l'*University of North Carolina* de 1933 à 1960 (1, 38);
- Frank **Yates** (1902-1994), à Rothamsted de 1931 à 1968 (13, 18);
- et William Gemmell **Cochran** (1909-1980), à Rothamsted de 1934 à 1939 et à l'*Iowa State College* de 1939 à 1957 (39).

J'ai en fait déjà débordé ainsi dans la seconde moitié du vingtième siècle, mais je voudrais encore associer, aux noms que j'ai cités, celui de David John **Finney** (1917-), qui a oeuvré à Rothamsted de 1939 à 1945 et a enseigné à l'Université d'Aberdeen et à l'Université d'Edimbourg de 1954 à 1984, en dirigeant l'Unité de Statistique de l'*Agricultural Research Council* (25).

Huit noms et, au total, près de 250 années d'activités liées à l'agronomie!

Mais le mouvement n'est pas à sens unique. Il ne s'agit pas seulement de huit scientifiques qui mettent leurs compétences au service de l'enseignement et de la recherche agronomiques. Il s'agit aussi de huit scientifiques qui, partant très souvent de l'agronomie au sens large, développent des concepts nouveaux, utilisés ensuite dans de nombreuses autres disciplines: distributions t (de Student), inférence statistique relative aux moyennes et aux coefficients de corrélation, distributions F (de Fisher-Snedecor), analyse de la variance et de la covariance, échantillonnage, plans expérimentaux, analyse discriminante, etc.

Il s'agit encore des auteurs de remarquables livres de référence, dont certains ont d'ailleurs été traduits en plusieurs langues, et notamment en français. On peut citer, principalement: *Statistical methods for research workers* et *The design of experiments* de **Fisher** (15, 16), *Sampling techniques* de **Cochran** (7), *Experimental designs* de **Cochran** et **Cox** (8), *Statistical methods* de **Snedecor**, puis **Snedecor** et **Cochran** (32, 34), et *Sampling techniques for censuses and surveys* de **Yates** (36, 37).

Enfin, on retrouve aussi, parmi les personnes citées, les fondateurs de la Société internationale de Biométrie (en 1947) et de la revue *Biometrics* (en 1948), et plusieurs anciens présidents et vice-présidents de cette société.

À posteriori, on peut se demander ce qui a suscité des liens aussi étroits entre l'agronomie et la statistique ou la biométrie. Un élément important me semble être le fait que les agronomes

ont été parmi les premiers à devoir maîtriser la variabilité parfois considérable de leur matériel expérimental, végétal ou animal, et à devoir prendre, malgré cette variabilité, des décisions qui engagent l'avenir, souvent à long terme.

C'est en effet au dix-neuvième siècle qu'ont été organisées, en matière de fumure minérale et de rotation des cultures, les premières expériences de longue durée et que la variabilité importante des résultats obtenus est apparue, des conseils devant néanmoins être donnés aux agriculteurs. L'accumulation des résultats des expériences organisées à Rothamsted à partir du milieu du dix-neuvième siècle fut d'ailleurs une des raisons de l'engagement de **Fisher** et l'analyse de ces résultats y fut sa première tâche (6).

En particulier, les forestiers n'ont pas attendu la biométrie et les outils de 1920 pour concevoir les moyens de gestion que sont les tables de production: les premiers travaux réalisés dans ce domaine datent de la fin du dix-neuvième et du tout début du vingtième siècle (14, 24).

Les principes de base de la construction de ces tables sont très semblables à ceux qui ont été mis en oeuvre, à partir de 1930 seulement, en ce qui concerne les modèles économétriques (22, 23). Et de même, il a fallu attendre la Seconde Guerre mondiale pour que le secteur industriel s'intéresse réellement, sous la pression des militaires, à la gestion ou la maîtrise de la qualité, déjà développée dans les années 1920 et 1930 (12, 21), ainsi qu'à la planification des expériences et à l'analyse de la variance.

4. LA SECONDE MOITIÉ DU VINGTIÈME SIÈCLE: L'IMPACT DE L'ORDINATEUR

La seconde moitié du vingtième siècle est essentiellement marquée par le développement fulgurant de nouveaux moyens de calcul. Les premiers ordinateurs commercialisés – Ferranti et Univac – datent en effet du tout début des années cinquante (27). Et la machine Elliott installée à Rothamsted en 1954 est sans doute le premier – ou un des premiers – ordinateur à fonctionner dans un centre de recherche agronomique (40).

Très schématiquement, on peut distinguer différentes étapes dans l'évolution de l'emploi de l'ordinateur en matière statistique (28, 35).

Dans un premier temps, l'ordinateur est considéré comme une machine à calculer plus perfectionnée que les précédentes, qui assure un traitement plus rapide et plus complet des données. Des méthodes statistiques tout à fait classiques, telle l'analyse de la variance, sont alors appliquées de façon plus efficace, l'ordinateur permettant d'effectuer très facilement des transformations de variables, des estimations de données manquantes et, aussi, un contrôle plus strict des conditions d'application. Au même moment, des méthodes anciennes, sous-utilisées en raison de l'importance des calculs qu'elles nécessitent, deviennent tout à fait courantes. Tel est le cas notamment pour nombre de méthodes d'analyse statistique à plusieurs variables.

Dans un deuxième temps, des méthodes anciennes sont l'objet de développements théoriques nouveaux, parfois très importants. On peut citer dans cette optique l'analyse factorielle des correspondances et la classification numérique, qui s'inscrivent dans la ligne de méthodes datant du début du siècle. On peut y associer aussi la mise au point d'algorithmes de calcul, dont l'emploi se substitue progressivement à la consultation de tables.

Dans un troisième temps, les possibilités offertes par l'ordinateur conduisent au développement de méthodes entièrement nouvelles, telles les méthodes de rééchantillonnage – *jackknife* et *bootstrap*. Mais on pourrait parler aussi des méthodes robustes, de simulation, de modélisation, etc.

Le recours à l'ordinateur s'accélère encore à partir des années quatre-vingt du fait de l'apparition, puis de la très large diffusion des micro-ordinateurs ou ordinateurs personnels (4), dont les capacités dépassent aujourd'hui très largement celles des "grands" systèmes des années

soixante. L'ordinateur est ainsi devenu un outil que chacun peut utiliser à tout moment à sa meilleure convenance – du moins dans certaines parties du Monde, car nous oublions trop souvent qu'un nombre important de nos collègues des pays en voie de développement ne disposent pas des mêmes facilités que nous.

Plus récemment encore, viennent les réseaux internationaux de transmission de données – Internet ou autres – qui assurent des communications directes et instantanées entre les scientifiques des différents pays.

J'ai ainsi parlé essentiellement, jusqu'à présent, des matériels informatiques et des infrastructures. Je ne peux cependant terminer sans souligner l'importance considérable – pour ne pas dire vitale – des logiciels, en particulier statistiques. D'abord peu nombreux et relativement généraux, d'un emploi qui n'était pas toujours des plus faciles et essentiellement orientés vers le calcul, ces logiciels sont progressivement devenus très diversifiés et parfois très spécialisés, extrêmement conviviaux et orientés aussi vers d'autres fonctions, telles que la présentation graphique des données (2, 20).

Je ferai remarquer enfin qu'en ce qui concerne la seconde moitié du vingtième siècle, je n'ai pratiquement pas évoqué l'agronomie. Je pense en effet que l'évolution de la statistique et de la biométrie a été beaucoup plus intégrée que précédemment et n'a plus été liée aussi étroitement qu'auparavant à tel ou tel secteur d'application.

5. CONCLUSION

En résumé, l'examen du passé met en évidence le développement parallèle de l'agronomie et de la statistique, ainsi que la naissance de la biométrie au dix-neuvième siècle, des relations particulièrement étroites entre agronomie, statistique et biométrie au cours de la première moitié du vingtième siècle, et l'influence prépondérante de l'informatique, et en particulier de la micro-informatique, durant la seconde moitié du vingtième siècle.

Je m'efforcerai de vous donner aussi une certaine vision de l'avenir, quand nous aurons entendu Elisabeth de **Turckheim** et Jean **Tranchefort**.

RÉFÉRENCES BIBLIOGRAPHIQUES

- (1) ANDERSON R.L., MONROE R.J., NELSON L.A., 1979 – Gertrude M. Cox: a modern pioneer in statistics. *Biometrics* **35** (1) 3-7.
- (2) BABILLIOT A., 1988 – Typologie des progiciels statistiques pour micro-ordinateur. *Rev. Stat. Appl.* **36** (3), 15-32.
- (3) BARTLETT M.S., 1956 – John Wishart, D.Sc., F.R.S.E. *J. R. Stat. Soc., Ser. A*, **119** (4), 492-493.
- (4) BOARDMAN T.J., 1982 – The future of statistical computing on desktop computers. *Amer. Stat.* **36** (1), 49-58.
- (5) BOLAND P.J., 1984 – A biographical glimpse of William Sealy Gosset. *Amer. Stat.* **38** (3), 179-183.
- (6) BOX J.F., 1978 – *R.A. Fisher, the life of a scientist*. New York, Wiley, 512 p.
- (7) COCHRAN W.G., 1953 – *Sampling techniques*. New York, Wiley, 330 p.
- (8) COCHRAN W.G., COX G.M., 1950 – *Experimental designs*. New York, Wiley, 454 p.
- (9) COX G.M., HOMEYER P.G., 1975 – Professional and personal glimpses of George W. Snedecor. *Biometrics* **31** (2), 265-301.
- (10) DAGNELIE P., 1992 – *Statistique théorique et appliquée. Tome 1: les bases théoriques*. Gembloux, Presses agronomiques, 502 p.
- (11) DROESBEKE J.J., TASSI P., 1990 – *Histoire de la statistique*. Paris, Presses universitaires de France, 128 p.
- (12) DUNCAN A.J., 1959 – *Quality control and industrial statistics*. Homewood, Irwin, 946 p.

- (13) DYKE G., 1995 – Obituary: Frank Yates. *J. R. Stat. Soc., Ser. A*, **158** (2), 333-338.
- (14) EICHHORN F., 1902 – *Ertragstafeln für die Weisstanne*. Berlin, Springer, 106 p.
- (15) FISHER R.A., 1925 – *Statistical methods for research workers*. Edinburgh, Oliver and Boyd.
- (16) FISHER R.A., 1935 – *The design of experiments*. Edinburgh, Oliver and Boyd.
- (17) HANKINS F.H., 1908 – *Adolphe Quetelet as statistician*. New York, Columbia University, 134 p.
- (18) HEALY M.J.R., 1995 – Frank Yates, 1902-1994. *Biometrics* **51** (1), 389-391.
- (19) IRWIN J.O., 1959 – Biometric method: past, present and future. *Biometrics* **15** (3), 363-375.
- (20) KOCH A., HAAG U. (ed.), 1995 – The statistical software guide '94/95. *Comput. Stat. Data Anal.* **19** (2), 237-261.
- (21) KOTZ S., JOHNSON N.L. (ed.), 1982-1988 – *Encyclopedia of statistical sciences* (9 vol.). New York, Wiley, 480 + 613 + 722 + 657 + 741 + 758 + 714 + 870 + 762 p.
- (22) LANGE O., 1970 – *Leçons d'économétrie* (trad. A. Posner). Paris, Gauthier-Villars, 425 p.
- (23) LESER C.E.V., 1968 – A survey of econometrics. *J. R. Stat. Soc., Ser. A*, **131** (4), 530-566.
- (24) LOREY T., 1899 – *Ertragstafeln für die Fichte*. Frankfurt, Sauerländers.
- (25) MACNEILL I., 1993 – A conversation with David J. Finney. *Stat. Sci.* **8** (2), 187-201.
- (26) MAILLY E., 1875 – *Essai sur la vie et les ouvrages de L.A.J. Quetelet*. Bruxelles, Hayez, 191 p.
- (27) MOREAU R., 1981 – *Ainsi naquit l'informatique: les hommes, les matériels à l'origine des concepts de l'informatique d'aujourd'hui*. Paris, Dunod, 222 p.
- (28) MULLER M.E., 1970 – Computers as an instrument for data analysis. *Technometrics* **12** (2), 259-293.
- (29) PLACKETT R.L., BARNARD G.A. (ed.), 1990 – *'Student': a statistical biography of William Sealy Gosset based on writings by E.S. Pearson*. Oxford, Clarendon, 142 p.
- (30) QUETELET A., 1846 – *Lettres à S.A.R. le duc régnant de Saxe-Cobourg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politiques*. Bruxelles, Hayez, 450 p.
- (31) RASCH D., TIKU M.L., SUMPFF D. (ed.), 1994 – *Elsevier's dictionary of biometry in English, French, Spanish, Dutch, German, Italian and Russian*. Amsterdam, Elsevier, 887 p.
- (32) SNEDECOR G.W., 1937 – *Statistical methods applied to experiments in agriculture and biology*. Ames, Iowa State College Press.
- (33) SNEDECOR G.W., 1954 – Biometry, its makers and concepts. In: Kempthorne O., Bancroft T.A., Gowen J.W., Lush J.L. (ed.). *Statistics and mathematics in biology*. Ames, Iowa State College Press, 3-11.
- (34) SNEDECOR G.W., COCHRAN W.G., 1971 – *Méthodes statistiques* (trad. H. Boelle et E. Camhaji). Paris, Association de Coordination technique agricole, 649 p.
- (35) VICTOR N., 1984 – Computational statistics: tool or science (with discussion). *Stat. Softw. Newsl.* **10** (3), 105-125.
- (36) YATES F., 1949 – *Sampling techniques for censuses and surveys*. London, Griffin, 318 p.
- (37) YATES F., 1951 – *Méthodes de sondage pour recensements et enquêtes* (trad. G. Darmois). Paris, Dunod, 335 p.
- (38) YATES F., 1979 – Gertrude Mary Cox, 1900-1978. *J. R. Stat. Soc., Ser. A*, **142** (4), 516-517.
- (39) YATES F., 1982 – William Gemmell Cochran, 1909-1980. *J. R. Stat. Soc., Ser. A*, **145** (4), 521-523.
- (40) YATES F., REES D.H., 1958 – The use of an electronic computer in research statistics: four year's experience. *Comp. J.* **1**, 49-58.
- (41) X, 1979 – *Hommage à Jean Dufrenoy, 1894-1972*. Paris, Académie d'Agriculture de France, 188 p.