

COLLABORATION ENTRE L'OFFICE FÉDÉRAL DE LA STATISTIQUE ET L'UNIVERSITÉ DE NEUCHÂTEL : BILAN ET PERSPECTIVES

Yves Tillé¹

¹ *Université de Neuchâtel, Suisse, yves.tille@unine.ch*

Résumé. En 2001, l'Office fédéral de la statistique et l'Université de Neuchâtel ont signé un accord visant à promouvoir la recherche dans le domaine de la statistique officielle. L'accord prévoyait que des recherches sur des sujets d'intérêt pour l'Office fédéral de la statistique soient menées à l'université. Depuis 24 ans, cette collaboration est positive pour les deux parties. D'une part, l'Université de Neuchâtel a obtenu des ressources supplémentaires pour développer la recherche fondamentale, notamment par le biais de la formation doctorale. D'autre part, l'Office fédéral de la statistique a bénéficié directement des retombées de cette recherche. Un grand nombre de résultats ont été publiés dans des revues scientifiques grâce à cette collaboration. Nous passons en revue les principaux résultats dans les domaines de l'échantillonnage, de la coordination des échantillons, de l'estimation de la variance, du calibrage et du traitement de la non-réponse.

Mots-clés. formation doctorale, recherche fondamentale, statistique officielle.

Abstract. In 2001, the Swiss Federal Statistical Office and the University of Neuchâtel signed an agreement to promote research in the field of official statistics. The agreement provided for research on topics of interest to the Federal Statistical Office to be conducted at the university. For 24 years, this collaboration has been positive for both parties. On the one hand, the University of Neuchâtel has obtained additional resources to develop fundamental research, in particular through doctoral training. On the other hand, the Federal Statistical Office has benefited directly from the spin-offs of this research. A large number of results have been published in scientific journals thanks to this collaboration. We review the main results in the areas of sampling, sample coordination, variance estimation, calibration and the treatment of non-response.

Keywords. doctoral training, basic research, official statistics.

1 Collaboration Université de Neuchâtel-Office fédéral de la statistique

L'année 1998 marque un tournant pour l'Office fédéral de la statistique (OFS), avec le regroupement de l'ensemble de ses services à Neuchâtel, achevant ainsi le processus de

décentralisation depuis Berne. L'OFS se situe face à la gare de Neuchâtel. La collaboration avec l'Université de Neuchâtel a débuté en 2001, coïncidant avec mon arrivée dans la ville. Précédemment, j'avais eu l'honneur de travailler pendant six ans à l'ENSAI à Rennes, où ma coopération avec Jean-Claude Deville s'était révélée particulièrement fructueuse.

L'objectif principal était de nouer une collaboration en matière de recherche. L'OFS était particulièrement intéressé par le développement de recherches dans les domaines stratégiques de la statistique officielle. Ces travaux ont été menés à bien à l'Université de Neuchâtel. Le succès de cette collaboration réside probablement dans le fait que chaque institution est restée fidèle à sa mission première. L'OFS cherchait à améliorer la qualité de sa production statistique tout en bénéficiant de la proximité d'une équipe de recherche dynamique dans ses domaines d'intérêt. De son côté, l'université a pu accroître ses ressources pour mener des recherches fondamentales. De nombreux sujets ont été abordés, dont une liste est présentée ci-dessous.

2 Thèmes de recherche

2.1 Échantillonnage à probabilités inégales

Un premier thème abordé concerne l'échantillonnage à probabilité inégale. À première vue, ce sujet peut sembler anecdotique, mais il s'avère en réalité captivant. On peut se demander pourquoi des centaines, voire des milliers, de publications sont dédiées à ce domaine. Une des raisons est que l'implémentation de plans d'échantillonnage à probabilités inégales sans remise et de taille fixe est un problème à la fois complexe et contre-intuitif.

Pour s'en convaincre, il suffit d'utiliser la fonction `sample()` dans R pour constater qu'elle produit des résultats erronés. Ci-dessous, les probabilités d'inclusion ont été simulées en utilisant cette fonction. En calculant les statistiques de test pour une comparaison des proportions, on constate clairement que `sample()` ne respecte pas les probabilités d'inclusion théoriquement attendues.

Le code suivant permet de calculer par simulations

$$\frac{\sqrt{n}(\hat{\pi}_k - \pi_k)}{\sqrt{\pi_k(1 - \pi_k)}},$$

où π_k est la probabilités prescrite et $\hat{\pi}_k$ est la probabilité obtenue.

```

install.packages("sampling")
library(sampling)
N=10;n=5
pik=inclusionprobabilities(1:N,n)
pik
S=rep(0,N)
SIM=10000
for(i in 1:SIM)
{
s=rep(0,N)
s[sample.int( N,n, replace = FALSE, prob = pik/n)]=1
S=S+s
}
pike=S/SIM
pik
(pike-pik)/sqrt(pik*(1-pik)/SIM)

```

Dans le domaine des recherches sur les données massives, l'échantillonnage est redevenu un sujet d'actualité. On y a recours principalement parce qu'il est souvent impossible de stocker l'intégralité des données. Cependant, on observe que les mêmes erreurs commises par les statisticiens il y a plus de 50 ans, avant que ces problématiques ne soient clarifiées, sont simplement répétées. Nous avons proposé de multiples innovations dans ce domaine (Matei et Tillé, 2005a, 2006, 2007; Pea *et al.*, 2007; Qualité, 2008; Haziza et Tillé, 2010; Grafström *et al.*, 2012; Tillé, 2016c; Tillé *et al.*, 2018b; Wilhelm *et al.*, 2017; Tillé, 2019a; Panahbehagh *et al.*, 2023a; Jauslin *et al.*, 2022; Panahbehagh *et al.*, 2023b; Tillé, 2023; Tillé et Panahbehagh, 2024).

2.2 Échantillonnage équilibré

L'échantillonnage équilibré consiste à sélectionner dans un registre un ou plusieurs échantillons dont les estimateurs des totaux sont égaux ou presque égaux aux totaux de la population. Le développement de l'échantillonnage équilibré par la méthode du cube a été l'innovation principale résultant de ma collaboration avec Jean-Claude Deville. Cependant, la publication de cet article a été un long parcours semé d'embûches, et il n'a finalement été publié qu'en 2004 après de nombreuses difficultés.

Plusieurs questions étaient restées en suspens. Comment implémenter efficacement l'algorithme ? Dans quel ordre trier les unités ? Quelles sont les probabilités d'inclusion optimales ? Étant donné que le complément d'un échantillon équilibré n'est pas forcément équilibré, comment peut-on compléter ou coordonner un tel échantillon de manière appropriée ? Toutes ces questions ont été abordées (Deville et Tillé, 2004, 2005; Tillé et Favre, 2004; Chauvet et Tillé, 2006a, 2007; Tillé et Favre, 2005; Tillé, 2011; Vallée *et al.*, 2015; Marazzi et Tillé, 2017; Leuenberger *et al.*, 2022; Tillé et Rivest, 2024; Tripet et Tillé, 2024).

2.3 Réflexions sur l’optimisation des plans

L’optimisation des plans d’échantillonnage est une question cruciale. Nous avons montré que sous un modèle linéaire un échantillon équilibré avec des probabilités d’inclusion inégales proportionnelles aux écarts types des termes d’erreur est le plan optimal. Sous un modèle linéaire avec des conditions simples, l’estimateur optimal dans le cadre de l’inférence basée sur le modèle est égal à l’estimateur de Horvitz-Thompson si le plan est optimal. Ces résultats fournissent une méthodologie pour concevoir des plans appropriés (Nedyalkova et Tillé, 2008a, 2012; Tillé et Wilhelm, 2017).

2.4 Échantillonnage spatial

Dans l’espace, les données sont souvent autocorrélées. Si on sélectionne des unités proches les unes des autres, on obtient alors moins d’informations. Pour estimer un total, il est alors préférable d’étaler l’échantillon. Il est également possible d’étaler un échantillon sur un espace de variables auxiliaires, ce qui produira un effet de stratification locale généralisée. Des progrès significatifs ont été réalisés, mais le problème n’est pas complètement résolu dans le sens où il n’existe pas de méthode donnant les meilleurs plans étalés pour les cas particuliers où les plans sont directement identifiables, comme par exemple le plan systématique à deux dimensions (Grafström et Tillé, 2013; Tillé et Ecker, 2013; Dickson et Tillé, 2016; Tillé *et al.*, 2018a; Jauslin et Tillé, 2020a; Ecker *et al.*, 2023).

2.5 Coordination des échantillons

La coordination des échantillons vise à gérer la rotation et le recouvrement entre une suite d’échantillons. Par exemple, dans certaines enquêtes, 25% des individus sont remplacés chaque trimestre. Ce problème peut sembler inextricable, car il est impossible d’optimiser simultanément les plans transversaux et la coordination entre les échantillons. Théoriquement, cela se traduit par une suite d’échantillons représentée sous forme de matrice d’indicatrices, où chaque indicatrice représente la sélection d’une unité à un moment donné. Lorsqu’on suit une unité dans le temps, on obtient un échantillon longitudinal pour cette unité. Dans le cas d’une coordination négative, le meilleur échantillon longitudinal est systématique, ce qui signifie qu’après quelques vagues, la sélection devient déterministe. À ce stade, il n’est plus possible de contrôler efficacement le plan transversal. Il faut donc faire un choix entre l’optimisation du plan transversal et celle du plan longitudinal. À l’OFS, le choix a été de privilégier le plan longitudinal afin de garantir une rotation optimale. Les échantillons transversaux, quant à eux, sont sélectionnés au moyen d’un plan de Poisson (Matei et Tillé, 2004, 2005b; Matei, 2008; Nedyalkova et Tillé, 2008b; Nedyalkova *et al.*, 2008, 2009; Matei et Skinner, 2009; Nedyalkova *et al.*, 2011; Graf et Qualité, 2014; Grafström et Matei, 2015; Matei et Grafström, 2016),

2.6 Calage

Le calage est une méthode incontournable utilisée par tous les statisticiens d'enquête. Bien que de nombreuses nouvelles techniques aient été proposées, beaucoup d'entre elles produisent des poids qui dépendent de la variable d'intérêt. Nous avons eu l'opportunité d'apporter des améliorations à ces méthodes. Par exemple, en imposant des bornes sur les poids, en calculant des bornes minimales, ou en utilisant le calage "mou" lorsque le nombre de variables est très élevé. On a également proposé des méthodes d'harmonisation de poids quand des variables ont des poids différents (Tillé, 2002; Berger *et al.*, 2003; Graf, 2014b; Devaud et Tillé, 2019a,b; Guandalini et Tillé, 2017; Guggemos et Tillé, 2010; Tripet et Tillé, 2023).

2.7 Indices d'inégalités de revenus et leur variance

Un enjeu important de la statistique publique est le suivi des conditions de vie des ménages. La mesure des inégalités salariales en constitue un aspect essentiel. Nous avons concentré nos efforts sur l'inférence statistique relative à ces indices. Ce problème est complexe et présente de nombreuses facettes, notamment en raison de la difficulté à modéliser les revenus (Langel et Tillé, 2013, 2011b,a, 2012; Tillé et Langel, 2012; Graf et Tillé, 2014b,a; Graf et Nedyalkova, 2014a, 2017, 2014b; Dong *et al.*, 2024, 2021).

2.8 Écart salarial entre femmes et hommes

Un autre enjeu important concerne l'écart salarial entre les hommes et les femmes. La méthode classique, proposée par Alan Blinder et Ronald Oaxaca, consiste à établir deux modèles de régression distincts pour prédire les logarithmes des revenus, l'un pour les hommes et l'autre pour les femmes. Ensuite, il est possible d'appliquer le modèle des femmes aux hommes et inversement, afin de décomposer l'écart salarial en une partie explicable et une partie non explicable.

Cependant, cette méthodologie présente des limites, notamment en raison de la difficulté à modéliser les salaires. En effet, les salaires ne suivent pas nécessairement une distribution log-normale. Nous avons démontré qu'au lieu de se concentrer uniquement sur la modélisation, il est possible de repondérer les observations. Par ailleurs, des modèles plus sophistiqués peuvent être utilisés, comme une distribution bêta généralisée de type 2. (Anastasiade Guinand *et al.*, 2023; Anastasiade et Tillé, 2017a,b).

2.9 Petits domaines

Des milliers d'articles ont été publiés sur le sujet de l'estimation pour les petits domaines. Le problème consiste à proposer des estimations pour des domaines pour lesquels peu ou pas de données sont disponibles. On utilise alors un modèle et des variables auxiliaires pour proposer des estimations. Cependant, il est surprenant de constater le faible nombre d'implémentations pérennes dans la production statistique. Ces méthodes présentent

pourtant un intérêt considérable, notamment pour déterminer l’attribution des subventions (comme celles de la Banque mondiale ou des fonds structurels de l’UE) et pour établir des cartes de pauvreté à un niveau de détail élevé. Ici encore, nous sommes confrontés à un problème de modélisation des revenus. Nous avons proposé des méthodes qui favorisent la cohérence des estimations obtenues (Graf *et al.*, 2019; Guadamarra *et al.*, 2020; Randrianasolo et Tillé, 2013).

2.10 Imputation par donneur

De nombreux articles sont consacrés à l’imputation. Dans la pratique, cependant, la situation est beaucoup plus complexe qu’un simple problème académique. La non-réponse peut apparaître dans toutes les variables et la manière d’aborder le problème n’est souvent pas claire. Il existe des méthodes itératives où chaque variable est imputée séquentiellement comme IVEware ou MissForest. Nous avons proposé plusieurs méthodes alternatives qui privilégient le respect des règles logiques par rapport à l’imputation (Favre *et al.*, 2004, 2005; Matei et Ranalli, 2015; Hasler et Tillé, 2014, 2016; Jauslin *et al.*, 2021b; Tripet *et al.*, 2023; Eustache *et al.*, 2024).

2.11 Estimation de variance

L’estimation de la précision constitue un véritable défi en théorie des sondages. En effet, la variance doit intégrer plusieurs éléments, notamment le plan de sondage, la non-réponse et son traitement, le calage, ainsi que l’éventuelle non-linéarité des estimateurs. Dans le cas des enquêtes répétées, il est également essentiel de prendre en compte la coordination entre les échantillons (Qualité et Tillé, 2008; Tillé et Vallée, 2017; Vallée et Tillé, 2019). Lorsque l’on souhaite estimer la variance par bootstrap, la tâche devient complexe, car en sondage, un estimateur de variance ne correspond pas directement à une variance. Une méthode bootstrap appliquée à un plan complexe ne peut pas simplement imiter le plan original. Il est donc nécessaire de construire un plan bootstrap spécifique pour chaque plan de sondage original afin d’obtenir des estimations correctes (Antal et Tillé, 2014, 2011a,b).

2.12 Autres travaux

D’autres sujets ont été abordés comme l’appariement d’échantillons qui peut être utilisé pour réaliser des estimations sur base de sources multiples (Jauslin et Tillé, 2023). Plusieurs ouvrages de synthèse ou de réflexion générale ont également été publiés (Berger et Tillé, 2009; Tillé, 2016a,b; Tillé et Matei, 2016; Tillé, 2022b,a; Ardilly *et al.*, 2023; Tillé, 2024) ainsi que des livres ou des monographies (Tillé *et al.*, 2020; Tillé, 2020, 2019b, 2010, 2006; Ardilly et Tillé, 2006, 2003; Tillé, 2001). Le développement de nombreux logiciels permettent d’utiliser directement les méthodes proposées (Eustache *et al.*, 2020; Eustache, 2020; Jauslin et Tillé, 2020b; Jauslin *et al.*, 2021a; Tillé et Matei, 2021; Chauvet et Tillé, 2006b). Enfin, la collaboration entre l’OFS et l’Université de Neuchâtel a contribué à plusieurs thèses de doctorat (Matei, 2005; Nedyalkova, 2009; Qualité, 2009; Langel, 2012; Antal, 2013; Randrianasolo,

2013; Graf, 2014a; Hasler, 2015; Wilhelm, 2017; Anastasiade, 2018; Vallée, 2019; Jauslin, 2023; Eustache, 2023; Dong, 2024).

3 L’avenir de la statistique officielle

Comment envisager l’avenir ? Une statistique officielle indépendante et neutre est un pilier essentiel de la démocratie et est d’une importance cruciale à une époque où la désinformation est omniprésente. Cette indépendance est fréquemment attaquée, notamment par des partis politiques extrémistes qui rejettent les institutions autonomes et impartiales.

L’évolution de la statistique officielle suit une trajectoire relativement claire. Partout, les taux de réponse aux enquêtes sont en déclin. Il devient donc essentiel, d’une part, de valoriser les sources administratives et, d’autre part, d’adopter des méthodes statistiques modernes pour optimiser la production statistique. Cette situation représente un véritable défi d’avenir, marquant le début d’une transition statistique incontournable.

Cependant, ces transformations rencontrent des difficultés. Les sources administratives dépendent étroitement de l’organisation administrative propre à chaque pays, ce qui complique l’harmonisation internationale. De plus, les données administratives ne sont souvent pas conçues pour la production statistique, ce qui rend leur valorisation extrêmement complexe.

L’adoption de nouvelles méthodes statistiques pose également des défis fondamentaux. La Charte de Statistique publique de la Suisse stipule que “Les informations statistiques sont documentées afin de faciliter leur compréhension et de permettre leur utilisation correcte.” “Les utilisateurs sont tenus informés de la méthodologie utilisée et de la qualité des résultats statistiques.” Avec une méthode d’apprentissage statistique, est-ce qu’une méthode avec laquelle on prédit sans comprendre est compatible avec l’éthique de la statistique publique ? Est-ce que les nouvelles méthodes statistiques sont compatibles avec le devoir de transparence ?

Pour relever ces défis, l’innovation est indispensable, et celle-ci nécessite un investissement significatif dans la recherche. Une synergie entre les universités et les instituts de statistique constitue une approche efficace pour développer cette recherche. Un aspect crucial de la recherche universitaire réside dans la formation doctorale, qui combine les missions de formation et de recherche scientifique. Cette formation doctorale permet non seulement de proposer des innovations méthodologiques, mais aussi de former des experts capables de répondre aux exigences de la transition statistique. Une collaboration avec l’université présente également l’avantage majeur de faciliter la diffusion des idées, car le milieu académique accorde une grande importance à la publication des résultats de recherche.

Bibliographie

ANASTASIADÉ, M. C. (2018). *Estimation of a counterfactual wage distribution using survey data*. Thèse de doctorat, Université de Neuchâtel.

- ANASTASIADÉ, M. C. et TILLÉ, Y. (2017a). Decomposition of gender wage inequalities through calibration: Application to the Swiss structure of earnings survey. *Survey Methodology*, 43:211–234.
- ANASTASIADÉ, M. C. et TILLÉ, Y. (2017b). Gender wage inequalities in Switzerland: the public versus the private sector. *Statistical Methods and Applications*, 26(2):293–316.
- ANASTASIADÉ GUINAND, M. C., MATEI, A. et TILLÉ, Y. (2023). Gender wage difference estimation at quantile levels using sample survey data. *TEST*, 32:1392–1433.
- ANTAL, E. (2013). *Variance estimation via resampling methods for complex sampling designs*. Thèse de doctorat, Université de Neuchâtel.
- ANTAL, E. et TILLÉ, Y. (2011a). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, 106:534–543.
- ANTAL, E. et TILLÉ, Y. (2011b). Simple random sampling with over-replacement. *Journal of Statistical Planning and Inference*, 141:597–601.
- ANTAL, E. et TILLÉ, Y. (2014). A new resampling method for sampling designs without replacement: the doubled half bootstrap. *Computational Statistics*, 29(5):1345–1363.
- ARDILLY, P., HAZIZA, D., LAVALLÉE, P. et TILLÉ, Y. (2023). Jean-Claude Deville’s contributions to survey theory and official statistics. *Survey Methodology*, 42:257–298.
- ARDILLY, P. et TILLÉ, Y. (2003). *Exercices corrigés de méthodes de sondage*. Ellipses, Paris.
- ARDILLY, P. et TILLÉ, Y. (2006). *Sampling Methods: Exercises and Solutions*. Springer, New York.
- BERGER, Y. G., EL HAJ TIRARI, M. et TILLÉ, Y. (2003). Toward optimal regression estimation in sample surveys. *Australian and New Zealand Journal of Statistics*, 45:319–329.
- BERGER, Y. G. et TILLÉ, Y. (2009). Sampling with unequal probabilities. In PFEFFERMANN, D. et RAO, C. R., éditeurs : *Handbook of Statistics. Vol 29A, Sampling*, pages 37–52. Elsevier, New York.
- CHAUVET, G. et TILLÉ, Y. (2006a). A fast algorithm of balanced sampling. *Journal of Computational Statistics*, 21:9–31.
- CHAUVET, G. et TILLÉ, Y. (2006b). Fastcube SAS-IML Macro. Université de Neuchâtel.
- CHAUVET, G. et TILLÉ, Y. (2007). Application of the fast SAS macros for balancing samples to the selection of addresses. *Case Studies in Business, Industry and Government Statistics*, 1:173–182.
- DEVAUD, D. et TILLÉ, Y. (2019a). Deville and Särndal’s calibration: revisiting a 25 years old successful optimization problem. *TEST*, 4:1033–1065.
- DEVAUD, D. et TILLÉ, Y. (2019b). Rejoinder on: Deville and Särndal’s calibration: revisiting a 25 years old successful optimization problem. *TEST*, 28:1087–1091.
- DEVILLE, J.-C. et TILLÉ, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91:893–912.
- DEVILLE, J.-C. et TILLÉ, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128:569–591.
- DICKSON, M. M. et TILLÉ, Y. (2016). Ordered spatial sampling by means of the traveling salesman problem. *Computational Statistics*, 31(4):1359–1372.
- DONG, Z. (2024). *On Measuring Income Inequality*. Thèse de doctorat, Université de Neuchâtel.
- DONG, Z., TILLÉ, Y., GIORGI, G. M. et GUANDALINI, A. (2021). Linearization and variance

- estimation of the Bonferroni inequality index. *Journal of the Royal Statistical Society*, A184(3): 1008–1029.
- DONG, Z., TILLÉ, Y., GIORGI, G. M. et GUANDALINI, A. (2024). Generalised income inequality index. *International Statistical Review*, 92:87–105.
- ECKER, K., MEIER, E. et TILLÉ, Y. (2023). Integrating spatial and ecological information into comprehensive biodiversity monitoring on agricultural land. *Environmental Monitoring and Assessment*, 195:1–20.
- EUSTACHE, E. (2020). *SwissCheese package: Imputation for Swiss cheese nonresponse*. R package version beta.
- EUSTACHE, E. (2023). *Nonresponse in Sample Surveys: New Estimation and Inference Methods*. Thèse de doctorat, Université de Neuchâtel.
- EUSTACHE, E., JAUSLIN, R. et TILLÉ, Y. (2020). *SpotSampling: SPatial and Optimally Temporal (SPOT) Sampling*. R package version 0.1.0
<https://cran.r-project.org/web/packages/SpotSampling/index.html>.
- EUSTACHE, E., VALLÉE, A.-A. et TILLÉ, Y. (2024). Balanced donor imputation handling Swiss cheese nonresponse. *Statistica Sinica*, 34:637–655.
- FAVRE, A.-C., MATEI, A. et TILLÉ, Y. (2004). A variant of the Cox algorithm for the imputation of non-response of qualitative data. *Computational Statistics and Data Analysis*, 45:709–719.
- FAVRE, A.-C., MATEI, A. et TILLÉ, Y. (2005). Calibrated random imputation for qualitative data. *Journal of Statistical Planning and Inference*, 128(2):411–425.
- GRAF, É. (2014a). *Imputation of income variables in a survey context and estimation of variance for indicators of poverty and social exclusion*. Thèse de doctorat, Université de Neuchâtel.
- GRAF, É. et QUALITÉ, L. (2014). Sondage dans des registres de population et de ménages en Suisse : coordination d'échantillons, pondération et imputation. Rapport technique, Université de Neuchâtel et Office Fédéral de la Statistique, Suisse.
- GRAF, É. et TILLÉ, Y. (2014a). Estimation de variance par linéarisation pour des indices de pauvreté et d'exclusion sociale. *Techniques d'enquête*, 40(1):69–88.
- GRAF, É. et TILLÉ, Y. (2014b). Variance estimation using linearization for poverty and social exclusion indicators. *Survey Methodology*, 40(1):61–79.
- GRAF, M. (2014b). Tight calibration of survey weights. Université de Neuchâtel.
- GRAF, M., MARÍN, J. M. et MOLINA, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *TEST*, 28(2):565–597.
- GRAF, M. et NEDYALKOVA, D. (2014a). Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind. *Review of Income and Wealth*, 60(4):821–842.
- GRAF, M. et NEDYALKOVA, D. (2014b). Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind. *Review of Income and Wealth*, 60(4):821–842.
- GRAF, M. et NEDYALKOVA, D. (2017). Discretizing a compound distribution with application to categorical modelling. *Statistics: A Journal of Theoretical and Applied Statistics*, 51:685–710.
- GRAFSTRÖM, A. et MATEI, A. (2015). Coordination of conditional Poisson samples. *Journal of Official Statistics*, 31(4):649–672.

- GRAFSTRÖM, A., MATEI, A., QUALITÉ, L. et TILLÉ, Y. (2012). Size constrained unequal probability sampling with a non-integer sum of inclusion probabilities. *Electronic Journal of Statistics*, 6:1477–1489.
- GRAFSTRÖM, A. et TILLÉ, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 14(2):120–131.
- GUADAMARRA, M., MOLINA, I. et TILLÉ, Y. (2020). Small area estimation methods under cut-off sampling. *Survey Methodology*, 46:51–78.
- GUANDALINI, A. et TILLÉ, Y. (2017). Design-based estimators calibrated on estimated totals from multiple surveys. *International Statistical Review*, 85:250–269.
- GUGGEMOS, F. et TILLÉ, Y. (2010). Penalized calibration in survey sampling: Design-based estimation assisted by mixed models. *Journal of Statistical Planning and Inference*, 140(11):3199–3212.
- HASLER, C. (2015). *New methods to handle nonresponse in surveys*. Thèse de doctorat, Université de Neuchâtel.
- HASLER, C. et TILLÉ, Y. (2014). Fast balanced sampling for highly stratified population. *Computational Statistics and Data Analysis*, 74:81–94.
- HASLER, C. et TILLÉ, Y. (2016). Balanced k -nearest neighbor imputation. *Statistics*, 105:11–23.
- HAZIZA, D. et TILLÉ, Y. (2010). An interesting property of the entropy of some sampling designs. *Survey Methodology*, 36(2):229–231.
- JAUSLIN, R. (2023). *Spatially balanced sampling, stratification and statistical matching*. Thèse de doctorat, Université de Neuchâtel.
- JAUSLIN, R., EUSTACHE, E., PANAHBEHAGH, B. et TILLÉ, Y. (2021a). *StratifiedSampling: Different Methods for Stratified Sampling*. R Foundation for Statistical Computing, Vienna, Austria. R package version 0.3.0.
- JAUSLIN, R., EUSTACHE, E. et TILLÉ, Y. (2021b). Enhanced cube implementation for highly stratified population. *Japanese Journal of Statistics and Data Science*, 4:783–795.
- JAUSLIN, R., PANAHBEHAGH, B. et TILLÉ, Y. (2022). Sequential spatially balanced sampling. *Environmetrics*, 33:1–17.
- JAUSLIN, R. et TILLÉ, Y. (2020a). Spatial spread sampling using weakly associated vectors. *Journal of Agricultural, Biological and Environmental Statistics*, 25(3):431–451.
- JAUSLIN, R. et TILLÉ, Y. (2020b). *WaveSampling: Weakly Associated Vectors Sampling*. R package version 0.1.1 <http://CRAN.R-project.org/package=WaveSampling>.
- JAUSLIN, R. et TILLÉ, Y. (2023). An efficient approach for statistical matching of survey data through calibration, optimal transport and balanced sampling. *Journal of Statistical Planning and Inference*, 225:121–131.
- LANGEL, M. (2012). *Measuring inequality in finite population sampling*. Thèse de doctorat, Université de Neuchâtel.
- LANGEL, M. et TILLÉ, Y. (2011a). Corrado Gini, a pioneer in balanced sampling and inequality theory. *Metron*, 69:45–65.
- LANGEL, M. et TILLÉ, Y. (2011b). Statistical inference for the quintile share ratio. *Journal of Statistical Planning and Inference*, 141(8):2976–2985.
- LANGEL, M. et TILLÉ, Y. (2012). Inference by linearization for Zenga’s new inequality index: a comparison with the Gini index. *Metrika*, 75(8):1093–1110.

- LANGEL, M. et TILLÉ, Y. (2013). Variance estimation of the Gini index: Revisiting a result several times published. *Journal of the Royal Statistical Society*, A176(2):521–540.
- LEUENBERGER, M., EUSTACHE, E., JAUSLIN, R. et TILLÉ, Y. (2022). Balancing a sample almost perfectly. *Statistics and Probability Letters*, 180:109229, 1–5.
- MARAZZI, A. et TILLÉ, Y. (2017). Using past experience to optimize audit sampling design. *Review of Quantitative Finance and Accounting*, 49(2):435–462.
- MATEI, A. (2005). *Computational aspects of sample surveys*. Thèse de doctorat, Université de Neuchâtel.
- MATEI, A. (2008). La coordination des échantillons en utilisant la sélection contrôlée. In GUILBERT, P., HAZIZA, D., RUIZ-GAZEN, A. et TILLÉ, Y., éditeurs : *Méthodes de sondage : applications aux enquêtes longitudinales, à la santé et aux enquêtes électorales*, pages 305–309. Dunod, Paris.
- MATEI, A. et GRAFSTRÖM, A. (2016). Coordination of spatial samples. COMPSTAT.
- MATEI, A. et RANALLI, M. G. (2015). Dealing with non-ignorable nonresponse in survey sampling: A latent modeling approach. *Survey Methodology*, 41(1):145–164.
- MATEI, A. et SKINNER, C. J. (2009). Optimal sample coordination using controlled selection. *Journal of Statistical Planning and Inference*, 139(9):3112–3121.
- MATEI, A. et TILLÉ, Y. (2004). On the maximal sample coordination. In ANTOCH, J., éditeur : *Proceedings in Computational Statistics, COMPSTAT'04*, pages 1471–1480, New York. Physica-Verlag/Springer.
- MATEI, A. et TILLÉ, Y. (2005a). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21(4):543–570.
- MATEI, A. et TILLÉ, Y. (2005b). Maximal and minimal sample co-ordination. *Sankhyā*, 67:590–612.
- MATEI, A. et TILLÉ, Y. (2006). Aspects computationnels dans des plans ordonnés. In LAVALLÉE, P. et RIVEST, L.-P., éditeurs : *Méthodes d'enquêtes et sondages : Pratiques européenne et nord-américaine*, pages 423–429. Dunod, Paris.
- MATEI, A. et TILLÉ, Y. (2007). Computational aspects of order πps sampling schemes. *Computational Statistics and Data Analysis*, 51:3703–3717.
- NEDYALKOVA, D. (2009). *Evaluation and development of strategies for sample coordination and statistical inference in finite population sampling*. Thèse de doctorat, Université de Neuchâtel.
- NEDYALKOVA, D., PEA, J. et TILLÉ, Y. (2008). Sampling procedures for coordinating stratified samples: Methods based on microstrata. *International Statistical Review*, 76:368–386.
- NEDYALKOVA, D., QUALITÉ, L. et TILLÉ, Y. (2009). General framework for the rotation of units in repeated survey sampling. *Statistica Neerlandica*, 63:269–293.
- NEDYALKOVA, D., QUALITÉ, L. et TILLÉ, Y. (2011). Tirages coordonnés d'échantillons poissoniens. In TREMBLAY, M. E., LAVALLÉE, P. et EL HAJ TIRARI, M., éditeurs : *Pratiques et méthodes de sondage*, Pratiques et méthodes de sondage, pages 323–328. Dunod, Paris.
- NEDYALKOVA, D. et TILLÉ, Y. (2008a). Optimal sampling and estimation strategies under linear model. *Biometrika*, 95:521–537.
- NEDYALKOVA, D. et TILLÉ, Y. (2008b). Tirages coordonnés d'échantillons stratifiés : méthodes basées sur des microstrates. In GUILBERT, P., HAZIZA, D., RUIZ-GAZEN, A. et TILLÉ, Y., éditeurs : *Méthodes d'enquêtes : applications aux enquêtes longitudinales, à la santé et aux*

- enquêtes électorales*, pages 309–314. Dunod, Paris.
- NEDYALKOVA, D. et TILLÉ, Y. (2012). Bias robustness and efficiency in model-based inference. *Statistica Sinica*, 22:777–794.
- PANAHBEHAGH, B., JAUSLIN, R. et TILLÉ, Y. (2023a). A general stream sampling design. *Computational Statistics*, **accepted**:1–32.
- PANAHBEHAGH, B., TILLÉ, Y. et AZAD KHANZADI, A. (2023b). Sampling with adaptive drawing probabilities. *Survey Methodology*, 49(1):191–211.
- PEA, J., QUALITÉ, L. et TILLÉ, Y. (2007). Systematic sampling is a minimal support design. *Computational Statistics & Data Analysis*, 51:5591–5602.
- QUALITÉ, L. (2008). A comparison of conditional Poisson sampling versus unequal probability sampling with replacement. *Journal of Statistical Planning and Inference*, 138:1428–1432.
- QUALITÉ, L. (2009). *Unequal probability sampling and repeated surveys*. Thèse de doctorat, Université de Neuchâtel.
- QUALITÉ, L. et TILLÉ, Y. (2008). Variance estimation of changes in repeated surveys and its application to the Swiss survey of value added. *Survey Methodology*, 34:173–181.
- RANDRIANASOLO, T. (2013). *Inférence basée sur le plan pour l'estimation de petits domaines / Design-based inference for small area estimation*. Thèse de doctorat, Université de Neuchâtel & Université Paris-Est, Neuchâtel, Paris.
- RANDRIANASOLO, T. et TILLÉ, Y. (2013). Small area estimation by splitting the sampling weights. *Electronic Journal of Statistics*, 7:1835–1855.
- TILLÉ, Y. (2001). *Théorie des sondages: échantillonnage et estimation en populations finies*. Dunod, Paris.
- TILLÉ, Y. (2002). Unbiased calibrated estimation on a distribution function in simple random sampling. *Survey Methodology*, 28:77–85.
- TILLÉ, Y. (2006). *Sampling Algorithms*. Springer, New York.
- TILLÉ, Y. (2010). *Muestreo equilibrado eficiente: el método del cubo*. Instituto Vasco de Estadística, Seminario internacional de estadística n° 52, Vitoria-Gasteiz.
- TILLÉ, Y. (2011). Ten years of balanced sampling with the cube method: an appraisal. *Survey Methodology*, 37:215–226.
- TILLÉ, Y. (2016a). A few remarks on a small example by Jean-Claude Deville regarding non-ignorable non-response. *Survey Methodology*, 42:307–317.
- TILLÉ, Y. (2016b). The legacy of Corrado Gini in survey sampling and inequality theory. *Metron*, 74(2):167–174.
- TILLÉ, Y. (2016c). Unequal probability inverse sampling. *Survey Methodology*, 42:283–295.
- TILLÉ, Y. (2019a). A general result for selecting balanced unequal probability samples from a stream. *Information Processing Letters*, 152:1–6.
- TILLÉ, Y. (2019b). *Théorie des sondages : échantillonnage et estimation en population finie*. 2^e édition, Dunod, Paris.
- TILLÉ, Y. (2020). *Sampling and Estimation From Finite Populations*. Wiley, Hoboken.
- TILLÉ, Y. (2022a). Some solutions inspired by survey sampling theory to build effective clinical trials. *International Statistical Review*, 90(3):481–498.

- TILLÉ, Y. (2022b). Some thoughts on official statistics and its future. *Journal of Official Statistics*, 38:669–671.
- TILLÉ, Y. (2023). Remarks on some misconceptions about unequal probability sampling without replacement. *Computer Science Review*, 47:100533, 1–6.
- TILLÉ, Y. (2024). Yet another attempt to classify positive univariate probability distributions. *Austrian Journal of Statistics*, 53:87–101.
- TILLÉ, Y., DICKSON, M. M. et ESPA, G. (2020). *Elementi di campionamento e stima da popolazioni finite*. Pearson, Milano.
- TILLÉ, Y., DICKSON, M. M., ESPA, G. et GIULIANI, D. (2018a). Measuring the spatial balance of a sample: A new measure based on the Moran’s I index. *Spatial Statistics*, 23:182–192.
- TILLÉ, Y. et ECKER, K. (2013). Complex national sampling design for long-term monitoring of protected dry grasslands in Switzerland. *Environmental and Ecological Statistics*, 21:1–24.
- TILLÉ, Y. et FAVRE, A.-C. (2004). Coordination, combination and extension of optimal balanced samples. *Biometrika*, 91:913–927.
- TILLÉ, Y. et FAVRE, A.-C. (2005). Optimal allocation in balanced sampling. *Statistics and Probability Letters*, 74:31–37.
- TILLÉ, Y. et LANGEL, M. (2012). Histogram-based interpolation of the Lorenz curve and Gini index for grouped data. *The American Statistician*, 66(4):225–231.
- TILLÉ, Y. et MATEI, A. (2016). Basics of sampling for survey research. In *SAGE Handbook of Survey Methodology*, pages 311–328. Sage Publications, Thousand Oaks, California.
- TILLÉ, Y. et MATEI, A. (2021). *sampling: Survey Sampling*. R package version 2.9.
- TILLÉ, Y. et PANAHBEHAGH, B. (2024). Maximum entropy sampling by a Markov chain process. *Journal of Survey Statistics and Methodology*, 12:232–248.
- TILLÉ, Y. et RIVEST, L.-P. (2024). Existence and applications of samples that are exactly balanced. Institut de Statistique, Université de Neuchâtel.
- TILLÉ, Y. et VALLÉE, A.-A. (2017). Revisiting variance decomposition when independent samples intersect. *Statistics and Probability Letters*, 130:71–75.
- TILLÉ, Y. et WILHELM, M. (2017). Probability sampling designs: Balancing and principles for choice of design. *Statistical Science*, 32(2):176–189.
- TILLÉ, Y., WILHELM, M. et QUALITÉ, L. (2018b). Sampling designs from finite populations with spreading control parameters. *Statistica Sinica*, 28:471–504.
- TRIPET, A., EUSTACHE, E. et TILLÉ, Y. (2023). Improving donor imputation using the prediction power of random forest: a combination of Swisscheese and Missforest. *Journal of Survey Statistics and Methodology*, **accepted**:1–16.
- TRIPET, A. et TILLÉ, Y. (2023). Analysis of survey data when each variable has its own weighting system. Institut de Statistique, Université de Neuchâtel.
- TRIPET, A. et TILLÉ, Y. (2024). Balanced sampling with inequalities: Application to category bounding, matrix rounding, and spread sampling. Institut de Statistique, Université de Neuchâtel.
- VALLÉE, A.-A. (2019). *Handling auxiliary variables in survey sampling and nonresponse*. Thèse de doctorat, Université de Neuchâtel.
- VALLÉE, A.-A., FERLAND-RAYMOND, B., RIVEST, L.-P. et TILLÉ, Y. (2015). Incorporating spatial and operational constraints in the sampling designs for forest inventories. *Environmetrics*, 26(8):

557–570.

VALLÉE, A.-A. et TILLÉ, Y. (2019). Linearization for variance estimation by means of sampling indicators: application to nonresponse. *International Statistical Review*, 87(2):347–367.

WILHELM, M. (2017). *Sampling with repulsion*. Thèse de doctorat, Université de Neuchâtel.

WILHELM, M., QUALITÉ, L. et TILLÉ, Y. (2017). Quasi-systematic sampling from a continuous population. *Computational Statistics and Data Analysis*, 105:11–23.