

Application de la loi de Benford au contrôle de qualité des modèles numériques de terrain

■ Laurent POLIDORI - Mhamad EL HAGE

La loi de Benford fait le constat empirique d'une régularité dans la distribution statistique du premier chiffre dans de nombreuses séries de nombres (géographie, sport, économie, etc.). Elle a été utilisée pour détecter des fraudes comptables ou électorales. Dans le même esprit, nous avons cherché à l'utiliser comme critère de vraisemblance pour évaluer la qualité des modèles numériques de terrain. Les métriques considérées sont l'altitude, la pente et l'ordre de Strahler.

MOTS-CLÉS

Loi de Benford, MNT, validation

Parmi les données disponibles dans un système d'information géographique, le modèle numérique de terrain (MNT) qui décrit la géométrie de la surface topographique, est particulièrement exigeant en termes de qualité. En effet, il est utilisé pour des applications très diverses, dont il faut concilier les exigences. De plus, il représente une surface familière qui s'offre quotidiennement à la vue, ce qui peut rendre inacceptable le moindre défaut. La plupart des opérations de contrôle visant à évaluer la qualité d'un MNT sont basées sur la comparaison des altitudes du MNT avec des points de contrôle au sol dont l'altitude est censée être connue avec une meilleure précision que celle du MNT. Le problème est que la précision de l'élévation n'est pas le critère de qualité le plus pertinent pour un certain nombre d'applications qui nécessitent souvent des informations sur les pentes et les formes du relief plutôt que sur les positions absolues (Polidori 1995). Dans des travaux antérieurs, nous avons ainsi mis en évidence la possibilité de contrôler également le respect de règles de vraisemblance (El Hage 2012, El Hage et al. 2012, Polidori et al. 2014).

La loi de Benford, qui décrit empiriquement la distribution statistique du premier chiffre dans des séries de nombres, ayant été utilisée comme critère de vraisemblance pour détecter des fraudes comptables ou électorales,

nous nous sommes interrogés sur la possibilité de l'appliquer au relief terrestre, pour évaluer la vraisemblance des modèles numériques de terrain.

La loi de Benford

Dans de nombreux domaines des sciences mais aussi de l'économie, du sport et bien d'autres, on constate dans les séries de nombres que la fréquence d'apparition de N comme premier chiffre décroît rapidement en fonction de N, en suivant approximativement une loi $\log(1+1/N)$, avec 1 dans 30 % des cas, 2 dans 18 % des cas et ainsi de suite (figure 1).

Ce constat empirique est la loi de Benford, du nom du physicien américain Frank Benford qui avait constaté

que les pages d'une table de logarithmes n'avaient pas le même degré d'usure, découverte déjà faite auparavant mais qu'il avait généralisée à d'autres domaines (Delahaye 2007). Cette loi difficilement explicable et pas toujours respectée met en évidence des régularités surprenantes. On retrouve ainsi la loi de Benford dans certaines suites mathématiques. C'est le cas de la célèbre suite de Fibonacci, qui commence par exemple par deux fois le nombre 1, et dont chacun des termes suivants est la somme des deux précédents : 1, 1, 2, 3, 5, 8, 13, etc. Le premier chiffre obéit presque parfaitement à la loi de Benford, une régularité qu'on ne retrouve pas pour le deuxième chiffre (cf. figure 2).

On la retrouve dans un certain nombre de variables naturelles issues du monde biologique ou géophysique, notamment des données géographiques comme les populations des villes ou la longueur des cours d'eau. La loi de Benford semble se vérifier surtout lorsque l'amplitude de variation couvre plusieurs ordres de grandeur, tandis qu'elle ne s'applique évidemment pas lorsque l'amplitude

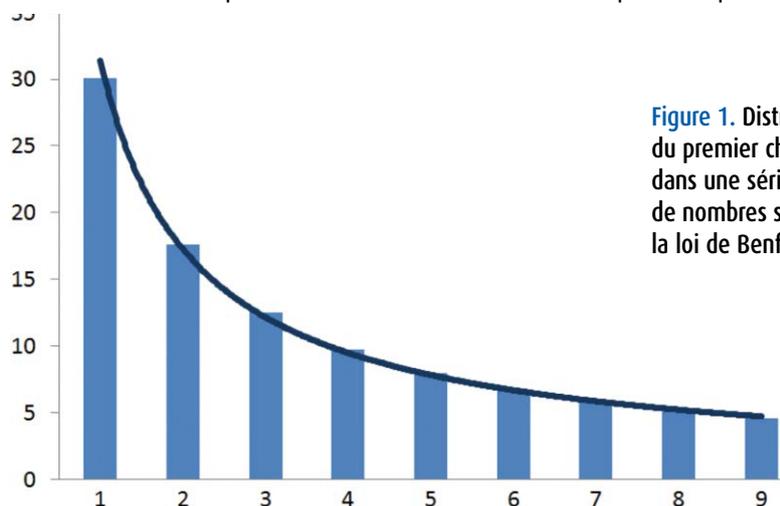


Figure 1. Distribution du premier chiffre dans une série de nombres selon la loi de Benford

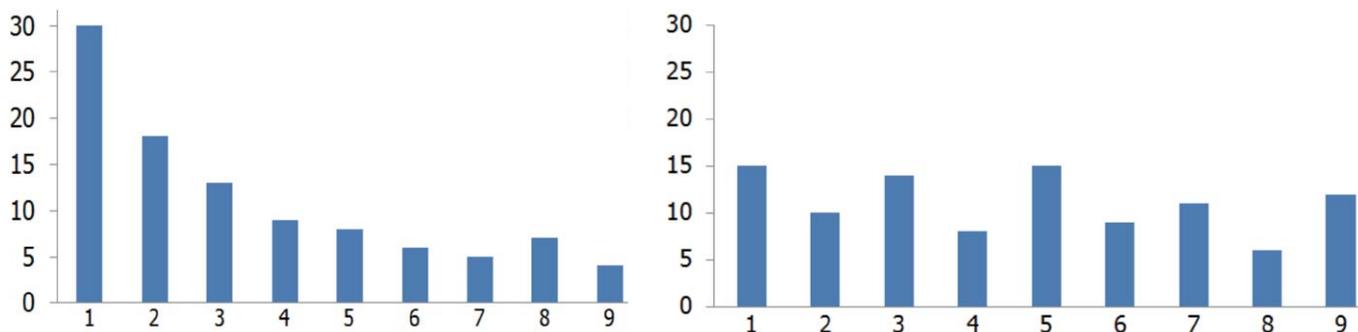


Figure 2. Distribution du premier chiffre (à gauche) et du deuxième chiffre (à droite) dans les 100 premiers termes de la suite de Fibonacci



de variation est naturellement limitée (comme la taille des hommes adultes en mètres, dont le premier chiffre est presque toujours 1, ou la superficie des départements français qui est presque toujours proche de 6 000 km² à la suite d'un découpage planifié du territoire).

Dans ces deux exemples, on peut s'interroger sur l'influence de l'unité choisie (ici m et km²) puisque la valeur numérique en dépend. En réalité, la loi de Benford est invariante par changement d'échelle. Un changement d'unité n'étant rien d'autre qu'un changement d'échelle, le changement d'unité est donc sans effet sur le respect ou non de la loi de Benford.

Recherche de lois logarithmiques dans le relief terrestre

Parmi les variables naturelles pour lesquelles Benford a vérifié le respect de la loi qui porte son nom, on trouve par exemple la longueur des rivières. Cela suggère que le relief terrestre présente des régularités selon la loi de Benford. Les rivières les plus longues sont les moins nombreuses, et cette raréfaction suit une loi logarithmique. Cela n'est pas sans rappeler la propriété d'autosimilarité de la côte bretonne, premier objet étudié par Mandelbrot (1975) pour illustrer la théorie des objets fractals : en mesurant la longueur de cette côte granitique avec une unité de longueur de plus en plus courte (ce qui revient à effectuer un zoom), on fait apparaître des détails de plus en plus petits et on trouve une distance de plus en plus grande, qui croît selon une loi logarithmique. Dans des études anté-

rieures, nous avons cherché à mettre en évidence des régularités dans le relief terrestre, soit pour mieux comprendre l'organisation spatiale de l'hydrographie (Polidori et al. 2014a), soit pour établir des règles de vraisemblance, afin de contrôler la qualité des MNT sans recourir à des points de contrôle (Polidori et al. 2014b). Pour mettre en évidence des régularités au sens de la loi de Benford, nous devons identifier des variables susceptibles d'être extraites d'un MNT et analyser la distribution statistique du premier chiffre, en espérant observer une décroissance logarithmique. Afin d'élargir les possibilités, nous ne nous limitons pas à la loi $\log(1+1/N)$ mais à toute forme de décroissance logarithmique fonction du premier

chiffre de la variable étudiée, ce qui nous permet de conserver l'esprit de la loi de Benford.

Nous avons effectué cette recherche sur un MNT extrait de la base altimétrique Topodata éditée par l'INPE au Brésil (Valeriano & Rossetti, 2012). La base Topodata est caractérisée par une maille de 30 m obtenue par densification du MNT SRTM initialement échantillonné à 90 m. La zone d'étude est située dans le Sud-Est du Brésil et présente un relief varié. Des tests supplémentaires ont été également effectués sur un MNT LiDAR à très haute résolution sur le Gabon.

L'altitude du MNT Topodata varie de 0 à 2635 m. Avec plusieurs ordres de grandeurs emboîtés, il n'est pas

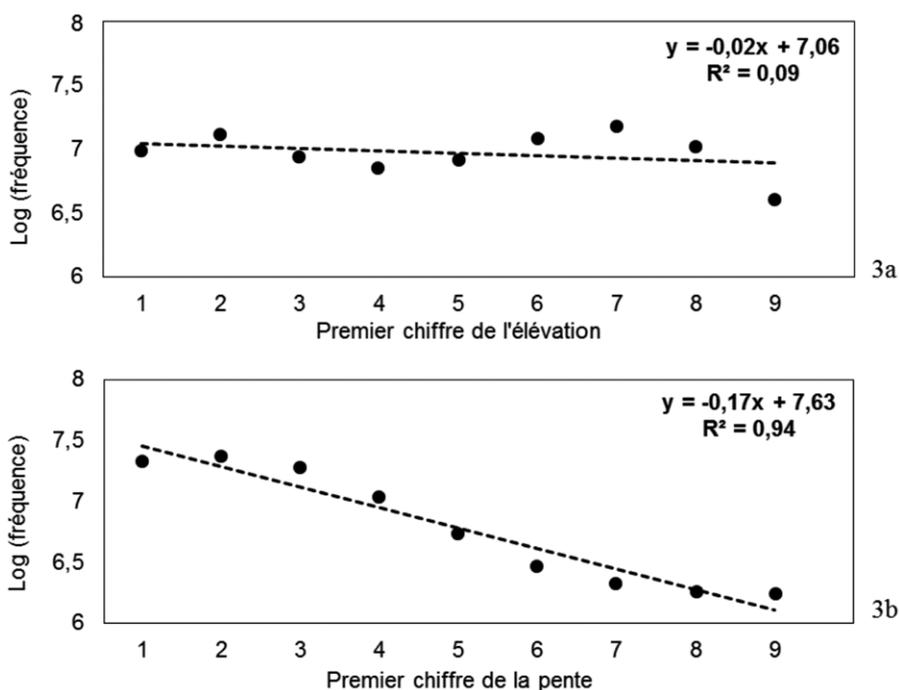


Figure 3. Distribution du premier chiffre des altitudes (3a) et du premier chiffre des pentes (3b) dans un MNT Topodata

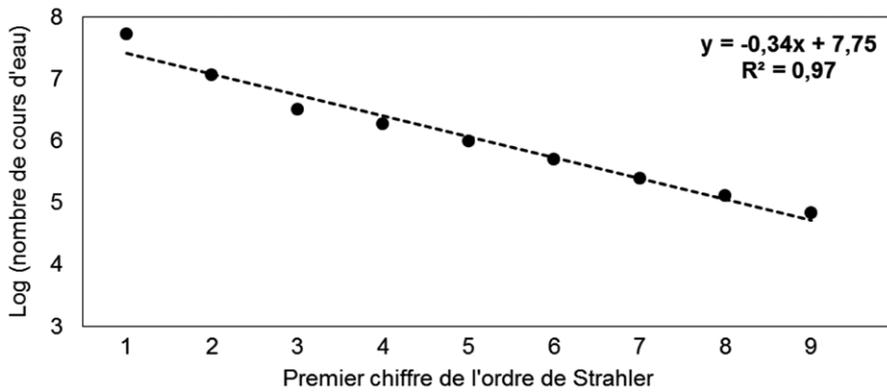


Figure 4. Nombre de tronçons de cours d'eau en fonction du premier chiffre de l'ordre de Strahler (MNT Topodata)

aberrant de se demander si l'altitude respecte la loi de Benford. Les résultats montrent qu'il n'en est rien : plus précisément, aucune régularité n'apparaît dans le premier chiffre de l'altitude (figure 3a). On observe une certaine régularité liée à la continuité de la variable d'altitude, mais aucune tendance logarithmique d'ensemble ($R^2=0,09$).

La pente varie de 0 à 527 pourcents (pourcentage défini par la tangente de l'angle entre le sol et le plan horizontal). Contrairement à ce que l'on observe pour l'altitude, le premier chiffre de la pente présente une distribution régulière avec une décroissance logarithmique caractérisée par $R^2=0,94$ (figure 3b). Des tests menés sur d'autres MNT mènent à des résultats similaires, jusqu'à $R^2=0,98$ pour un MNT LiDAR à très haute résolution (1 m) dans un bassin versant non anthropisé au Gabon.

La grandeur qui présente la tendance logarithmique la plus forte est l'ordre de Strahler. Dans un réseau hydrographique, chaque tronçon reliant une confluence amont et une confluence aval est caractérisé par un ordre dit de Strahler (Strahler 1952). Un tronçon reliant la source et la première confluence est d'ordre 1. Deux tronçons d'ordre 1 forment un tronçon d'ordre 2. Celui-ci ne change pas d'ordre lorsqu'il est rejoint par des tronçons d'ordre 1, mais sa confluence avec un autre tronçon d'ordre 2 donne un tronçon d'ordre 3, et ainsi de suite. Les tronçons de cours d'eau d'ordre 1 (de la source à la première confluence) sont plus nombreux que les tronçons

d'ordre 2, eux-mêmes plus nombreux que les tronçons d'ordre 3 et ainsi de suite. Cette raréfaction qui est conforme à l'intuition suit une loi logarithmique (figure 4) conséquente de la géométrie fractale du réseau hydrographique. Des tests menés sur de nombreux autres MNT conduisent au même résultat, c'est-à-dire à une linéarité presque parfaite.

Application au contrôle de qualité des MNT

La loi de Benford a été utilisée pour détecter des invraisemblances dans des domaines très éloignés de celui qui nous intéresse ici, notamment dans des bilans comptables, permettant de détecter des chiffres trafiqués (Nigrini, 2011). Cette loi empirique a pu permettre, non pas d'identifier avec certitude des erreurs dans des séries de chiffres mais de déclarer suspectes ces séries de chiffres. Dès lors, on est tenté de se demander si elle ne pourrait pas contribuer à décla-

rer suspectes des séries d'altitudes. L'idée est de déterminer une métrique qui a un comportement prévisible au sens de la loi de Benford sur la base d'observations répétées, à laquelle on puisse confronter les données réelles issues d'un MNT.

Si l'altitude est inopérante à offrir un tel critère de vraisemblance, on a vu qu'au contraire la pente est une métrique intéressante qui respecte la loi de Benford. La pente est d'ailleurs une métrique plus sensible que l'altitude aux erreurs du MNT, et la loi de Benford peut détecter une erreur de pente beaucoup plus sûrement qu'une erreur d'altitude. En effet, l'incertitude altimétrique va rarement affecter le premier chiffre de l'altitude sauf si un point à 1000 m, placé 1 m trop bas dans le MNT se retrouve à 999 m, ou si un point à 200 m est placé à 199 m, etc.

On peut donc se demander si le premier chiffre de la pente (ou plus exactement de la tangente de la pente exprimée en pourcents) respecte la loi de Benford, et interpréter un écart significatif à cette loi comme la manifestation d'une qualité dégradée. La figure 5 montre la distribution de la fréquence de répétition du premier chiffre de la pente (en pourcent) calculée à partir d'un MNT LiDAR très haute résolution (1 m) sur une zone urbanisée. Cette figure montre une régularité à partir de l'ordre 2 avec une valeur de R^2 presque égale à 1. Les pentes ayant 1 comme premier chiffre se trouvent un peu plus fréquentes, ce qui est dû à la présence des bâtiments qui font un angle droit avec les terrains avoisnants, et par conséquent, produisent

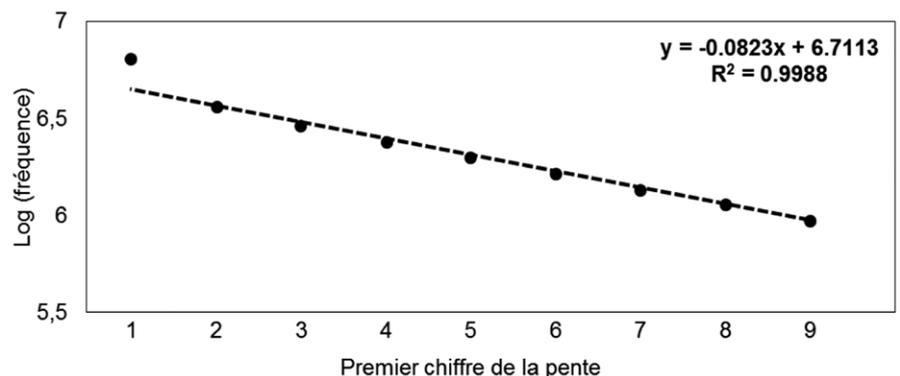


Figure 5. Mise en évidence d'une régularité logarithmique à partir de l'ordre 2 dans la distribution de la pente (MNT LiDAR à très haute résolution)

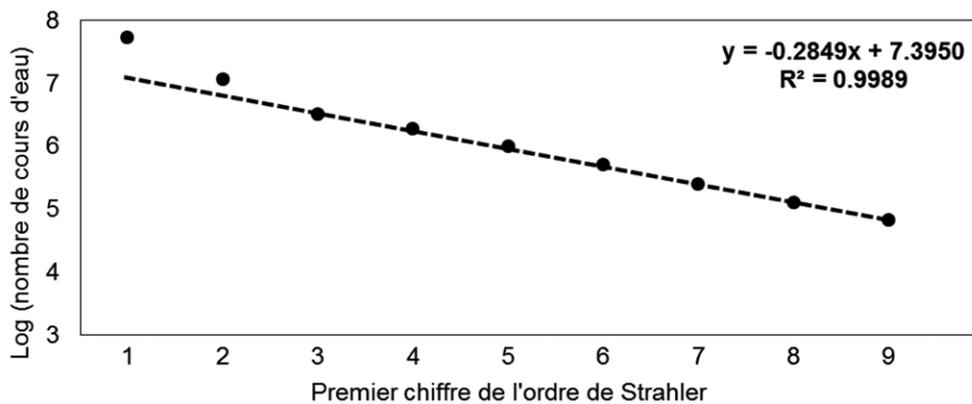


Figure 6. Mise en évidence d'une régularité logarithmique à partir de l'ordre 3 dans la distribution du nombre de cours d'eau en fonction du premier chiffre de l'ordre de Strahler (MNT Topodata)

► un nombre élevé de pentes supérieures à 1 000 % (> 84°). Un artefact dû à la présence d'objets artificiels est ainsi mis en évidence.

Le réseau hydrographique présente également presque toujours une régularité que nous avons décrite plus haut, et qui peut être vérifiée en examinant la variation du nombre total de tronçons de même ordre de Strahler en fonction de cet ordre. Pour les ordres les plus élevés, l'effectif réduit peut conduire à des statistiques peu représentatives et à un écart au modèle logarithmique que l'on ne peut pas imputer à la qualité du MNT. Au contraire, les ordres les plus petits (typiquement 1 et 2) correspondent à de petits cours d'eau qui sont plus difficiles à détecter et dont les motifs spatiaux s'expriment aux mêmes fréquences spatiales que les principaux artefacts du MNT (lissage, bruit, anisotropies liées à la structure raster...). La figure 6 reprend les mêmes données que la figure 4 en mettant en évidence un nombre anormalement élevé de tronçons d'ordre 2 et surtout 1, comparativement aux prédictions que l'on peut faire par un modèle logarithmique.

Conclusion

Nous avons présenté une application originale de la loi de Benford qui consiste à interpréter des régularités empiriques comme des critères de vraisemblance pour les modèles numériques de terrain afin d'identifier des défauts. Cette approche est justifiée par l'existence de régularités logarithmiques dans le premier chiffre de certaines variables qui caractérisent

le relief terrestre, notamment la pente et l'ordre de Strahler. D'autres métriques pourraient être étudiées. Bien que ces régularités ne puissent être établies que de manière empirique, le fait qu'elles soient observées dans un très grand nombre de jeux de données permet d'en faire des critères de vraisemblances utiles pour la validation des MNT. Cela vient renforcer la légitimité des méthodes de validation de MNT qui n'utilisent pas de données de référence. ●

Références

- Delahaye J. P. (2007) *L'étonnante loi de Benford*. Pour la Science, p. 90-95.
- El Hage M. (2012) *Étude de la qualité géomorphologique de modèles numériques de terrain issus de l'imagerie spatiale*. Thèse de doctorat, CNAM.
- El Hage M., Simonetto E., Faour G. & Polidori L. (2012) *Evaluation of elevation, slope and stream network quality of SPOT DEMs*. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Volume 1-2, pp. 63-67.
- Nigrini M. J. (2011) *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. Wiley, 480 p.
- Polidori L., Claden M., Frelat R., El Hage M., Bendraoua F., Doliscar G., Mondesir P., Piard B. & Gaucherel C. (2014a). *Elaboration du référentiel hydrographique d'Haïti à partir d'un MNT ASTER*. *Revue Française de Photogrammétrie et de Télédétection*, n°205, pp. 49-57.
- Polidori L., El Hage M. & Valeriano M.M. (2014b) *Digital elevation model validation with no ground control: application to the*

Topodata DEM in Brazil. *Bol. Ciênc. Geod.*, sec. Artigos, Curitiba, v. 20, no 2, p. 467-479.

Polidori L. (1995) *Réflexions sur la qualité des modèles numériques de terrain*. *Bulletin de la Société Française de Photogrammétrie et de Télédétection*, N°139, p. 10-19.

Strahler A.N. (1952) *Hypsometric (area-altitude) analysis of erosional topology*, *Geological Society of America Bulletin*, vol. 63, no 11, 1952, p. 1117-1142.

Valeriano M. M. & Rossetti D. F. (2012) *Topodata: Brazilian full coverage refinement of SRTM data*. *Applied Geography (Sevenoaks, England)*, 32, p. 300-309.

Contacts

Laurent POLIDORI

laurent.polidori@cesbio.cnes.fr
CESBIO, Université de Toulouse, CNES/
CNRS/IRD/UPS/INRA, Toulouse, France

Mhamad EL HAGE

mhamad.elhage@ul.edu.lb
Laboratoire d'Études Géospatiales (LEG),
Faculté de Génie, Université Libanaise,
Tripoli, Liban

ABSTRACT

Benford's law makes the empirical observation of a regularity in the statistical distribution of the first digit in many series of numbers (geography, sport, economy etc.). It has been used to detect accounting or electoral fraud. In the same spirit, we have tried to use it as a realism criterion for digital elevation model quality assessment. The metrics considered are altitude, slope and Strahler order.