

Comparer le paradigme spatial au spatio-temporel pour estimer l'évolution d'indicateurs socio-économiques à partir d'images satellites

Robin Jarry¹, Marc Chaumont^{1,2}, Laure Berti-Équille³, Gérard Subsol¹

¹ LIRMM, Univ. Montpellier, CNRS, Montpellier, France

² Université de Nîmes, France

³ ESPACE-DEV, Univ. Montpellier, IRD, UA, UG, UR, Montpellier, France

1 Introduction

Ces dernières années, beaucoup de travaux de recherche sur l'estimation de la pauvreté à partir d'images satellites ont été proposés (voir par exemple [1]). Les méthodes utilisant l'apprentissage profond donnent des résultats permettant d'estimer la pauvreté avec une précision correcte dans des pays où il est difficile de mener des enquêtes de terrain. Pour autant, des travaux récents observent des difficultés pour estimer l'évolution de la pauvreté, c'est-à-dire, estimer les variations de la pauvreté sur une période de temps donnée [2]. Les méthodes existantes ne reposent que sur des données spatiales, et ne sont pas adaptés à estimer des évolutions [3]. Une idée pour améliorer ces résultats serait de considérer la dépendance temporelle. Par exemple, dans [2], les auteurs notent une amélioration importante en estimant l'évolution de la pauvreté avec deux observations à deux dates différentes de la même zone géographique. Nous proposons d'étendre cette idée à des séries temporelles d'images satellites (SITS), qui sont des observations répétées d'une même zone géographique, à des dates variables. Cette idée s'appuie sur la réussite de l'utilisation des SITS dans d'autres domaines applicatifs [4]. Nous proposons d'évaluer si l'utilisation des SITS permet d'améliorer les estimations de l'évolution de la pauvreté.

2 Comparer un modèle spatial à un modèle spatio-temporel

Deux zones d'études : Zanzibar et Damas. Les zones d'études choisies sont exposées 2. La zone d'étude (1) est un voisinage large autour de Zanzibar, en Tanzanie, contenant, entre autres, Dar Es Salam, la capitale économique du pays. Cette zone est d'intérêt applicatif dans le cadre de ces recherches. La zone d'étude (2) est un voisinage de la ville de Damas en Syrie, avec une partie du Liban, notamment la ville de Beyrouth. Elle est sélectionnée, car elle contient à la fois des zones de guerres, sur lesquelles on peut observer des diminutions de l'intensité lumineuse nocturne (ILN), avec des zones qui ont suivi un développement économique normal. Nous étudierons ces zones de 2000 à 2020.

SITS & ILN. Nous avons choisi d'utiliser les images satellites multispectrales Landsat-7, parce qu'elles couvrent l'ensemble du globe avec une résolution spatiale et tem-

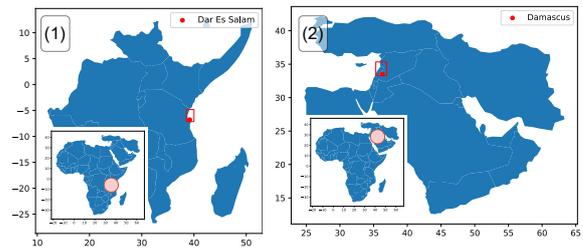


FIGURE 1 – La zone (1), à gauche, est un voisinage de Zanzibar. La zone (2), à droite, est un voisinage de Damas.

porable acceptable (30 mètres et 16 jours respectivement). Nous construisons à partir de ces observations des composites annuels, c'est-à-dire des observations moyennes (sans nuages) de notre zone d'étude pour chaque année. Nous avons choisi d'utiliser les intensités lumineuses nocturnes (ILN) de [5] comme données de référence à estimer, pour les mêmes raisons de couverture spatio-temporelle de Landsat-7, mais aussi, car les ILNs sont un proxy standard pour certains indicateurs socio-économiques [6]. Pour que la taille des images d'ILN de notre zone d'étude coïncide avec la taille des images Landsat-7, les ILNs sont sur-échantillonnées avec la méthode du plus proche voisin.

La base d'apprentissage. Le processus de construction du

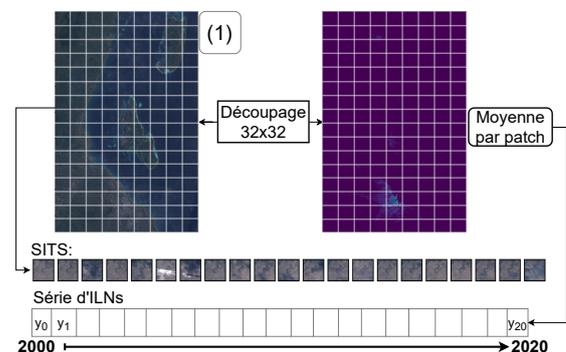


FIGURE 2 – Collecte des données et constitution du jeu d'apprentissage sur la zone (1). Le processus est identique sur la zone (2). Notons que les mailles de la grille sont plus petites en réalité.

Le jeu de données d'apprentissage est décrit en figure 2. Les SITS et séries d'ILN de la zone d'étude sont découpées en

Score		$MAE \downarrow (1)$	$R^2 \uparrow (1)$	$MAE \downarrow (2)$	$R^2 \uparrow (2)$
Par année	ST	0.085 \pm 0.010	0.695 \pm 0.063	1.011 \pm 0.061	0.543 \pm 0.014
	S	0.098 \pm 0.008	0.591 \pm 0.040	1.262 \pm 0.104	0.431 \pm 0.028
$\Delta t = 1$	ST	0.033 \pm 0.002	0.123 \pm 0.062	0.453 \pm 0.020	0.128 \pm 0.010
	S	0.087 \pm 0.009	-4.664 \pm 1.390	0.685 \pm 0.066	-1.240 \pm 0.439
$\Delta t = 10$	ST	0.103 \pm 0.005	0.322 \pm 0.042	1.083 \pm 0.043	0.218 \pm 0.013
	S	0.140 \pm 0.010	-0.291 \pm 0.083	1.287 \pm 0.056	-0.150 \pm 0.099
$\Delta t = 15$	ST	0.137 \pm 0.009	0.439 \pm 0.048	1.069 \pm 0.041	0.278 \pm 0.020
	S	0.163 \pm 0.010	-0.032 \pm 0.129	1.314 \pm 0.059	-0.123 \pm 0.071

TABLE 1 – Score MAE et R^2 sur les zones (1) et (2). Les exposants sont les écarts-types sur les cinq expériences.

patches, selon une grille régulière dont chaque maille fait 32×32 pixels. Les patches d’ILN de chaque pas de temps sont moyennés, résultant en une série de valeurs d’ILN sur la période temporelle d’étude. Un exemple d’apprentissage est alors une paire constituée d’une série de patches d’image Landsat-7 d’une même maille du quadrillage et l’évolution de l’ILN sur cette maille. Nos expériences sont menées avec une validation croisée à cinq plis. Il y a environ 40 000 exemples d’apprentissage dans chacune des zones (1) et (2).

Les architectures Transformer. Nous avons décidé de travailler avec l’architecture Transformer, car c’est une architecture de l’état de l’art ayant été adaptée à la fois pour des images, mais aussi des séquences d’images. Deux modèles sont construits, sur la base des travaux de [7] pour le modèle spatial (S), et des travaux de [4] pour le modèle spatio-temporel (ST). Le modèle spatial traite chaque image d’une séquence indépendamment des autres et estime une unique ILN par image. Nous l’utiliserons séquentiellement sur toutes les images d’une même SITS pour obtenir une prédiction d’évolution. Le modèle spatio-temporel analyse toutes les images de la SITS en une seule fois, recherchant des motifs spatiaux et temporels à corrélérer avec l’évolution à estimer.

3 Résultats

Par années. Dans la première ligne de score du tableau 1, nous calculons la moyenne des scores pour chacune des années de 2000 à 2020. Dans ce contexte, nous remarquons tout d’abord que les deux modèles atteignent des scores assez équivalents lorsqu’ils prédisent des valeurs d’ILN pour une année donnée, soit $R^2 = 0.69$ pour le modèle ST et $R^2 = 0.59$ pour le modèle S. Le score MAE est bien plus élevé sur la zone (2) que sur la zone (1), alors que les scores R^2 sont du même ordre (même si légèrement inférieurs sur la zone (2)), suggérant qu’il est plus difficile d’estimer l’ILN sur la zone (2) que sur la zone (1).

Par évolutions. Dans le reste du tableau 1, nous calculons la moyenne des scores sur les évolutions espacées de Δt années. Pour chaque exemple du JDD, l’évolution de l’ILN est obtenu en calculant la différence d’ILN à l’année $t + \Delta t$ et t . Nous observons que pour $\Delta t = 1$, *i.e.* une évolution d’un an, tous les scores R^2 sont soit négatifs, soit proche de 0 ce qui signifie que les dépendances à court terme sont difficiles à capturer dans les domaines spatial et temporel. Nous pensons que cela est dû au fait que les variations annuelles

de l’ILN sont difficilement perceptibles dans les données. Cependant, à mesure que Δt augmente, le modèle ST prédit mieux l’évolution avec $R^2 = 0,44$ pour $\Delta t = 15$ sur la zone (1). D’autre part, le modèle S reste peu fiable puisque $R^2 < 0$ pour tous les Δt . Individuellement, les scores R^2 des deux modèles s’améliorent lorsque Δt augmente. En revanche, ce dernier point n’est pas vérifié pour le score MAE, qui augmente dès que Δt augmente. En effet, les évolutions de l’ILN d’une année à l’autre n’ont pas la même amplitude que les évolutions de l’ILN sur une période de 10 ou 15 ans. Par conséquent, il est impossible de comparer les scores MAE pour des Δt différents.

4 Perspectives

Nous obtenons des résultats encourageants, qui nous permettent d’observer une supériorité du modèle spatio-temporel sur le modèle spatial. Néanmoins, nous avons fait notre analyse sur deux zones d’étude restreintes. De plus, les scores obtenus (notamment R^2) sont encore bien trop faibles pour une utilisation pratique de tels modèles. Une perspective est d’agrandir cette zone d’étude, à un continent entier par exemple. Cela permettrait de rendre notre analyse plus robuste et d’améliorer les performances des modèles.

Informations - Cet article est un résumé en français, avec des expériences complémentaires, de l’article suivant : R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. "Comparing Spatial and Spatio-Temporal Paradigms to Estimate The Evolution of Socio-Economical Indicators from Satellite Images." in *IGARSS 2023 - IEEE International Geoscience and Remote Sensing Symposium* (p. 5790-5793). 10.1109/IGARSS52108.2023.10282306

Références

- [1] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, August 2016.
- [2] C. Yeh, A. Perez, A. Driscoll, G. Azzari, Z. Tang, D. Lobell, S. Ermon, and M. Burke, "Using publicly available satellite imagery and deep learning to understand economic well-being in Africa," *Nature Communications*, vol. 11, no. 1, pp. 2583, May 2020.
- [3] L. Kondmann and X. X. Zhu, "Measuring changes in poverty with deep learning and satellite images," *ICLR, Practical ML for Developing Countries*, p. 6, 2020.
- [4] Y. Yuan, L. Lin, Q. Liu, R. Hang, and Z. Zhou, "SITS-Former : A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification," *Intl Journal of Applied Earth Observation and Geoinformation*, vol. 106, pp. 102651, Feb. 2022.
- [5] Z. Chen, B. Yu, C. Yang, Y. Zhou, S. Yao, X. Qian, C. Wang, B. Wu, and J. Wu, "An extended time series (2000–2018) of global NPP-VIIRS-like nighttime light data from a cross-sensor calibration," *Earth System Science Data*, vol. 13, no. 3, pp. 889–906, March 2021.
- [6] A. M. Noor, V. A. Alegana, P. W. Gething, A. J. Tatem, and R. W. Snow, "Using remotely sensed night-time light as a proxy for poverty in Africa," *Population Health Metrics*, vol. 6, no. 1, pp. 5, 2008.
- [7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale," in *ICLR*, 2021.