Hidden Markov Models for Annual Land Cover Mapping—Increasing Temporal Consistency and Completeness

Patrick Bogaert, Céline Lamarche, and Pierre Defourny, *Member, IEEE*

***Abstract*— This article aims at investigating the hidden Markov model (HMM) approach for the automated processing of classified satellite images for land cover and land-use change (LCLUC). HMM's account for transitions between classes at the same location, but that cannot be directly observed due to classification errors. Using a set of transition and emission probabilities, HMM's allow filtering out errors and recovering the actual sequence of LCLUC, which are typically overestimated when directly estimated from the classified images. After presenting the HMM framework, the methodology is illustrated on three 300-m annual time series of classified images from 2003 to 2019 over 756 × 756 km^2 areas in Brazil, People's Republic of China, and Mali. It is shown how the emission and transition probabilities can be estimated from these time series using a simple Viterbi training, alleviating computationally demanding algorithms. Special attention is paid to the processing of missing observations caused by clouds. Combining these three datasets with a simulation study, it is concluded that the HMM emission and transition probabilities can be estimated with low biases and variances thanks to the vast number (hundreds of thousands) of pixels at hand. The speed of the Viterbi training and decoding steps makes it possible to consider large-scale land cover mapping at moderate or even high spatial resolution as long as the legend of the LCLUC involves a reasonable number of classes like the six main Intergovernmental Panel on Climate Change (IPCC) land categories.**

***Index Terms*— Hidden Markov model (HMM), image classification, land cover, land-use change, remote sensing, Viterbi training.**

I. Introduction

LAND cover and land-use change (LCLUC) resulting from human activities has altered about 50% of Earth's land surface, primarily through natural vegetation conversion in the last century [1]. These changes have substantially affected Earth systems through altered biogeochemical and biogeophysical processes that change carbon, energy, and water fluxes to the atmosphere [2]. Monitoring the land cover dynamics at national to global scales is vital to quantify the current human footprint and the associated conversion rates in this global change context. Remote sensing provides comprehensive, wall-to-wall spatial coverage globally with imaging capacity and revisiting frequency allowing to detect rare or very localized landscape conversions like deforestation [3]–[5], urbanization [6], [7] and even highly dynamic transitions like shifting cultivation often overlooked by inventory-based statistics repeated with 3- to 5-year cycle [8], [9].

Land cover mapping mostly proceeds by labeling each pixel based on a classification process which is parametric or not and more recently based on machine learning or deep learning algorithms. Independent year-to-year classifications are known to suffer from spurious changes due to a combination of landscape inter-annual variability and poor discrimination between spectrally close land cover classes [10], [11]. Land cover change is known to be a rare event at a global scale [12]. Comparing annual land cover maps contaminated by classification errors to estimate land cover change leads to large overestimations. Land cover maps are nowadays typically produced annually for decades, sometimes using different sensors and classification algorithms, exacerbating key disagreements [13].

Several approaches have been proposed to deal with the temporal consistency issue of annual land cover map series. Desclee *et al* [14] relies on a multiyear segmentation to detect the forest change independently of classifications and related errors. For long-term time-series analysis [15], a set of descriptive change metrics were developed associated with various forest change events. In the Climate Change Initiative of the European Space Agency, the annual consistency of global annual land cover maps series 1992–2020 relies on the decoupling between change detection and land cover mapping [16], [17] to depict the land cover change based on the deviation of the observed versus the expected multispectral reflectance along the season. When comparing maps to assess land cover change, very few studies addressed the interweaving of mapping error and land cover change using a formal framework [18]. One solution could be the Hidden Markov Model (HMM), widely used in many fields [19] except multitemporal map comparison [13]. In our context, an HMM will account for spurious land cover change when comparing temporal sequences of land cover pixels impacted by classification errors. The HMM approach thus permits disentangling true land cover changes from classification errors inside the same framework.

In this article, we will show how the HMM can be effectively implemented over a large area to account for

Manuscript received May 11, 2021; revised September 6, 2021 and September 28, 2021; accepted October 12, 2021. Date of publication October 27, 2021; date of current version February 14, 2022. *(Corresponding author: Patrick Bogaert.)*

The authors are with the Earth and Life Institute Environmental Sciences, Université Catholique de Louvain (UCLouvain), 1348 Louvain-la-Neuve, Belgium (e-mail: patrick.bogaert@uclouvain.be).

Digital Object Identifier 10.1109/TGRS.2021.3123738

1558-0644 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

misclassification and missing data on the one hand, and on the other hand, to assess the actual changes that have occurred over time. Special attention will be given to implementing adaptive and automatic learning algorithms needed in a large-scale mapping application; simulations will evaluate their reliability. To illustrate the method's performance, three distinct areas widely distributed globally will be treated, each associated with a distinct amount of misclassifications and rates of change.

Abercrombie and Friedl [13] addressed the use of HMM to improve the consistency of annual land cover maps. Specifically, they integrated both the HMM approach and the classification procedure by starting from the spectral time series of the observations (which were processed with an ensemble decision tree classifier) to obtain the probability of classification errors used in the HMM. However, the rates of land cover change were selected somewhat arbitrarily for simplicity, and the goal was to obtain reliable maps of land cover at each time step in the sequence (based on a forward-backward algorithm and a maximum *a posteriori* probability criterion). In this article, however, we take a different approach by directly using the classified land cover maps and showing how the corresponding classification errors can be reliably rectified from these maps without prior information from the classification algorithm. This alleviates the complexity and computational burden of combining the classification algorithm and HMM estimation. We will focus on the most likely sequence of states using the Viterbi algorithm and Viterbi learning, leading to the most likely set of land cover maps over the entire temporal sequence. Our goal is to investigate the applicability of the HMM approach for large-scale consistent land cover mapping, focusing on machine and adaptive learning.

II. Material and Methods

A. Study Sites

Three sites of approximately 756 × 756 km^2 were selected to demonstrate the applicability of the HMM methodology in distinct landscapes affected by specific land cover transitions and different cloud coverage. For each site (in Brazil, the People's Republic of China (hereafter named China), and Mali), we will rely on yearly time series of 300-m surface reflectance images that combine Medium Resolution Imaging Spectrometer (MERIS) images from 2003 to 2011 and Project for On-Board Autonomy (PROBA-V) images from 2012 to 2019. Each year's seasonal mean composites were classified using a locally trained Gaussian maximum likelihood classifier implemented with a classification window of 100 × 100 pixels and a training window three times larger [20]. The European Space Agency Climate Change Initiative Land Cover (CCI-LC) map of 2015 with 22 LC classes was used as a training dataset for each study site. The classes from the resulting land cover maps were eventually aggregated in the six main Intergovernmental Panel on Climate Change (IPCC) land categories, namely Cropland, Forest, Grassland, Wetland, Settlement, and Others according to the rules described in [16]. Some pixels of the seasonal composites, flagged as clouds by the IdePix processor in the preprocessing steps, were not classified.

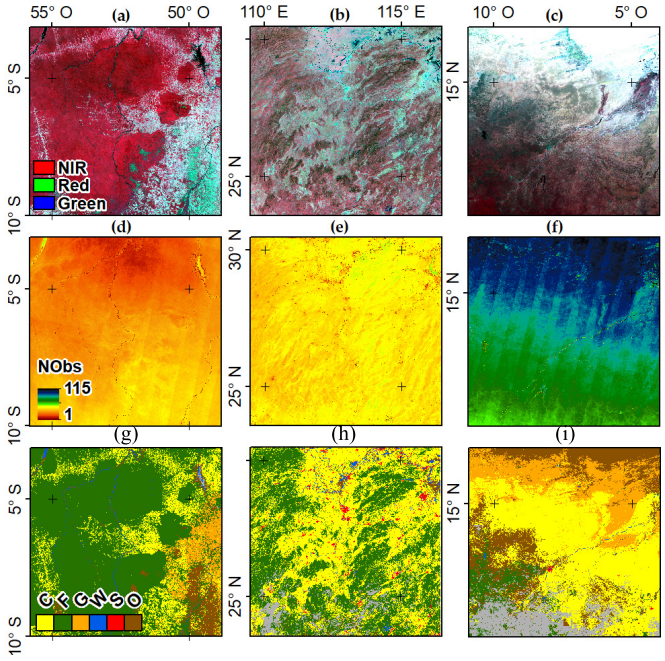


Fig. 1. (a)–(c) MERIS surface reflectance false-color composite over the three study areas in Brazil, China and Mali, respectively, averaged over the 2003–2011 period. (d)–(f) Corresponding maximum number of valid and cloud-free observations expected per year with a sensor revisiting capacity of five days. (g)–(i): Corresponding land cover maps for 2019 (the latest year in the temporal sequence), where each map covers 2520 × 2520 pixels at 300 m resolution. Labels are C = Cropland, F = Forest, G = Grassland, W = Water, S = Settlement, and O = Other. Gray pixels correspond to clouds.

The landscape of the Brazilian site (centered on 49.5°W, 4.5°S) is dominated by forests (70%) that are progressively converted to agricultural lands [8]. Most of the Chinese landscape (centered on 113°E, 27°N) is divided into croplands (50%) and forests (43%). Although water and settlements both cover only 2% of the surface, urban expansion and dams construction are frequent in China over these years [21]. The Malian site (centered on 7.5°S, 13.5°N) experienced the fewest land cover changes of all sites but presents high inter-annual land cover seasonality. Annual cloud coverage conditions highly vary across sites. On average, the maximum number of valid and cloud-free observations acquired per year at the five-day MERIS acquisition frequency were the fewest in Brazil (27), followed by China (40) and Mali (115). Fig. 1 is briefly presenting some of the properties associated with these three sites.

B. Hidden Markov Model

This section will briefly remind the fundamental ideas and results at the HMM root when applied to a set of annual image classification output in the framework of change detection and accuracy assessment. We will focus first on a single pixel that has been classified at equally space-time instants, with m possible classes. For two arbitrary time instants t and $t+1$, let us denote i_t and i_{t+1} as the true (but unknown) classes for that pixel and let us denote $\widehat{i}_t$ and $\widehat{i}_{t+1}$ as the labeled classes from a nonperfect classification procedure, where all these indices are taking their values in $\{1, \ldots, m\}$. Considering now jointly the time instants $t, t+1, \ldots, t+q$, what is sought for is

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

to recover the sequence of true classes $\mathbf{i} = (i_t, i_{t+1}, \ldots i_{t+q})$ knowing the labeled ones $\widehat{\mathbf{i}} = (\widehat{i}_t, \widehat{i}_{t+1}, \ldots \widehat{i}_{t+q})$. If it is assumed that probabilistic rules can capture the changes of the true classes over time and the classification errors, this can be modeled in the framework of an HMM using some additional assumptions. Accordingly, we will briefly remind in the next paragraphs the essential elements that are needed to explain the rationale of using an HMM for processing annual classified images, that can symmetrically be viewed for each pixel on these images as a sequence $\widehat{\mathbf{i}}$ of labeled classes over time. We will also provide detailed information about how the estimation of the HMM parameters can be learned from the data at hand. We will consider the possible absence of classification results due to, e.g., the presence of a cloud cover.

1) Maximum Likelihood Estimation: Focusing on a single pixel, it is possible to estimate the (nonobservable) true classes $\mathbf{i}$ from the set of labeled classes $\widehat{\mathbf{i}}$ by looking for $\mathbf{i}$ so that the probability $p(\mathbf{i}|\widehat{\mathbf{i}})$ is maximized. This is the classical maximum likelihood framework. Using basic probability calculus and few assumptions as detailed in the Appendix, the general formulation of an HMM model is then given by

$$p\left(\mathbf{i}|\widehat{\mathbf{i}}\right) \propto p(i_t) \prod_{j=1}^{q} p\left(i_{t+j}|i_{t+j-1}\right) \prod_{j=0}^{q} p\left(\widehat{i}_{t+j}|i_{t+j}\right) \quad (1)$$

where these probabilities do not depend on j for a stationary HMM (i.e. they are the same whatever the value for j). In that case, the HMM model refers to three sets of distinct probabilities, that are namely: 1) the initial probabilities $p(i_t)$; 2) the so-called transition probabilities $p(i_t|i_{t-1})$ that reflect the rate of change between two successive time instants; and 3) the so-called emission probabilities $p(\widehat{i}_t|i_t)$ that reflect the quality of the classification. The "hidden" part of the model comes from the fact that only the labeled classes $\widehat{\mathbf{i}}$ are at hand, while the true classes $\mathbf{i}$ are unknown but are linked to the labeled ones through the emission probabilities. A compact way to present these probabilities is through the $m \times 1$ vector $\mathbf{p}$ and the $m \times m$ $\mathbf{T}$ and $\mathbf{E}$ stochastic matrices, with

$$\begin{aligned} \mathbf{p} &= (p(1_t), \ldots, p(m_t))' \\ \mathbf{T} &= \begin{pmatrix} p(1_t|1_{t-1}) & \cdots & p(1_t|m_{t-1}) \\ p(2_t|1_{t-1}) & \cdots & p(2_t|m_{t-1}) \\ \vdots & \ddots & \vdots \\ p(m_t|1_{t-1}) & \cdots & p(m_t|m_{t-1}) \end{pmatrix} \\ \mathbf{E} &= \begin{pmatrix} p\left(\widehat{1}_t|1_t\right) & \cdots & p\left(\widehat{1}_t|m_t\right) \\ p\left(\widehat{2}_t|1_t\right) & \cdots & p\left(\widehat{2}_t|m_t\right) \\ \vdots & \ddots & \vdots \\ p\left(\widehat{m}_t|1_t\right) & \cdots & p\left(\widehat{m}_t|m_t\right) \end{pmatrix} \end{aligned} \quad (2)$$

where the sum of $\mathbf{p}$ and the sum of each column of $\mathbf{T}$ and $\mathbf{E}$ are equal to 1. In a remote-sensing framework, $\mathbf{T}$ and $\mathbf{E}$ matrices are related to distinct issues. The first is the change assessment through transition probabilities, while the second is the accuracy assessment through emission probabilities. Both problems are linked here together inside the HMM framework. As the amount of change is linked to specific processes that occur over limited areas in general, it is expected that $\mathbf{T}$ is diagonally dominant concerning its columns (i.e. the magnitude of the diagonal entry in a column is larger than or equal to the sum of the magnitudes of all the other entries in that column), with a high probability that i_{t+1} and i_t are identical. Similarly, as long as there are no major confusions between classes during the classification process and that the classification algorithm is reasonably efficient, it is expected that $\mathbf{E}$ is diagonally dominant, i.e. that producer's accuracies are above 0.5. However, none of these conditions is required in an HMM, and these conditions are not met in general for applications of the HMM in other contexts.

2) Viterbi Decoding: If we assume that $\mathbf{p}$, $\mathbf{T}$ and $\mathbf{E}$ are known, estimating the true classes $\mathbf{i}$ from the labeled ones $\widehat{\mathbf{i}}$ by maximizing $p(\mathbf{i}|\widehat{\mathbf{i}})$ is easily done. Indeed, maximizing $p(\mathbf{i}|\widehat{\mathbf{i}})$ in (1) with respect to $\mathbf{i}$ is equivalent to minimizing the negative log-likelihood

$$-\ln p\left(\mathbf{i}|\widehat{\mathbf{i}}\right) \propto -\ln p(i_t) - \sum_{j=1}^{q} \ln p\left(i_{t+j}|i_{t+j-1}\right) - \sum_{j=0}^{q} \ln p\left(\widehat{i}_{t+j}|i_{t+j}\right). \quad (3)$$

3) Viterbi Training: Let us consider that we focus on a classified area composed by a large set of n pixels that have been classified for the same $q+1$ time instants and let us denote $\{\widehat{\mathbf{i}}_1, \ldots, \widehat{\mathbf{i}}_n\}$ as the whole set of these labeled sequences. Stated in other words, we are considering the same subset of n pixels on classified maps at hand at $q+1$ equally spaced time instants. What is sought for is an estimate of the corresponding true classes $\{\mathbf{i}_1, \ldots, \mathbf{i}_n\}$ based on an HMM model, where the same parameters $\mathbf{p}$, $\mathbf{T}$, $\mathbf{E}$ are supposed to be relevant for all pixels but are however unknown.

Assuming $\mathbf{p}^{(0)}$, $\mathbf{T}^{(0)}$, $\mathbf{E}^{(0)}$ as initial guesses for these parameters, a first set of estimates $\{\mathbf{i}_1^{(0)}, \ldots, \mathbf{i}_n^{(0)}\}$ of the true classes can be obtained for each sequence by minimizing (3) using the Viterbi algorithm. From these true classes estimates (i.e., these Viterbi paths), one can in turn compute new estimates $\mathbf{p}^{(1)}$, $\mathbf{T}^{(1)}$, while from the pairs of sequences $(\widehat{\mathbf{i}}_1, \mathbf{i}_1^{(0)}), \ldots, (\widehat{\mathbf{i}}_n, \mathbf{i}_n^{(0)})$ a new estimate of $\mathbf{E}^{(1)}$ can be computed as well. The whole procedure can be repeated until convergence, i.e. until successive iterations lead to the same final estimate set. Because the new parameters of the HMM at each iteration are completely determined by the sequences of true classes estimates (i.e., the Viterbi paths) obtained from the Viterbi algorithm, successive iterations lead to identical estimates as soon as the Viterbi paths no longer change. Alternatively, the iterations can be stopped earlier as soon as a convergence criterion is satisfied, e.g., when the modifications of the HMM parameters' estimates between two iterations drop below a user-specified threshold.

In practice, the frequency estimates $f(\cdot)$ of all probabilities $p(\cdot)$ can be used, so that at the ℓth iteration we have

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

$\mathbf{p}^{(\ell)} = \{f^{(\ell)}(i_t)\}$, $\mathbf{T}^{(\ell)} = \{f^{(\ell)}(i_t|i_{t-1})\}$ and $\mathbf{E}^{(\ell)} = \{f^{(\ell)}(\widehat{i}_t|i_t)\}$, with

$$\begin{aligned} f^{(\ell)}(i_t) &= \frac{n^{(\ell)}(i_t)}{n} \\ f^{(\ell)}(i_t|i_{t-1}) &= \sum_{j=1}^{q} n^{(\ell)}\left(i_{t+j}, i_{t+j-1}\right) \Big/ \sum_{j=1}^{q} n^{(\ell)}\left(i_{t+j-1}\right) \\ f^{(\ell)}\left(\widehat{i}_t|i_t\right) &= \sum_{j=0}^{q} n^{(\ell)}\left(\widehat{i}_{t+j}, i_{t+j}\right) \Big/ \sum_{j=0}^{q} n^{(\ell)}\left(i_{t+j}\right) \end{aligned} \tag{4}$$

where $n^{(\ell)}(\cdot)$ are the counts for the corresponding (pairs of) classes at the ℓth iteration over the set of n pixels at the corresponding time instants. For the sake of notation simplicity, let us denote $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$, and $\mathbf{E}^{(\infty)}$ as the final estimate of the probabilities at the end of the iterative procedure, while $\mathbf{i}^{(\infty)}$ is the corresponding Viterbi path.

It is well documented that Viterbi training finds a local optimum of the likelihood by providing HMM parameters that maximize the contribution from the set of Viterbi paths to the likelihood and that this local optimum can be sensitive to the choice of the initial guesses. Based on several case studies that will be presented hereafter, we will show, however, that the biases of these estimated parameters are low in most instances, thanks to the large number n (typically millions) and to the sound initial guesses that can be made for $\mathbf{T}^{(0)}$, $\mathbf{E}^{(0)}$.

4) Case of Unchanged Pixels: In our specific framework, it is expected that the majority of pixels will not be impacted by any change over a relatively short time span so that for most pixels in our study areas, there is a single hidden class that characterizes the full sequence over time. For these pixels, changes for the attributed classes are thus only related to classification errors, and we have $i_t = i_{t+1} = \cdots = i_{t+q}$, that we can simply denote as i. Accordingly, the transition probabilities are no longer relevant, and the Viterbi decoding thus reduces to estimate i from the labeled sequence $\widehat{\mathbf{i}}$, so that (3) becomes

$$-\ln p\left(i|\widehat{\mathbf{i}}\right) \propto -\ln p(i) - \sum_{j=0}^{q} \ln p\left(\widehat{i}_{t+j}|i\right). \tag{5}$$

Finding i reverts to compute (5) for all m possible choices for the hidden class i and to select i such that (5) is minimized. This corresponds to a classical maximum likelihood estimation procedure from an i.i.d. sample $\widehat{\mathbf{i}}$, that does not involve any more a Markov model but that only depends on the emission probabilities $\mathbf{E}$. Clearly, the estimation quality will increase for each pixel with the sample size (i.e. sequence length). Moreover, from a Viterbi training algorithm applied on these pixels, it is clear too that the quality of $\mathbf{E}^{(\infty)}$ as an estimate of $\mathbf{E}$ will increase with the number of pixels involved in the training. Consequently, for a typical situation where millions of pixels are at hand for the Viterbi training, one expects to get low bias and variance for the $\mathbf{E}$ estimate. In practice, this would assume that we are first able to select pixels for which no change occurred to estimate $\mathbf{E}$ from these pixels, and afterward to use this $\mathbf{E}$ estimate in the HMM for estimating $\mathbf{T}$ using the remaining set of (changed) pixels. Unfortunately, there is no prior information about changed or unchanged pixels over the sequence, so this is impossible. However, if the fraction of changed pixels is low over the training area, this allows us to expect that $\mathbf{E}^{(\infty)}$ will be indeed a quite good estimate of $\mathbf{E}$. Indeed, if all Viterbi paths correspond to unchanged pixels so that $\mathbf{T}^{(\infty)} = \mathbf{I}$ (with $\mathbf{I}$ the identity matrix), then for each pixel, the estimated true class is precisely the class that maximizes (5). As a consequence, for areas where changes are limited so that $\mathbf{T}^{(\infty)} \simeq \mathbf{I}$, it is expected that $\mathbf{E}^{(\infty)}$ will have low variance and a limited bias.

5) Missing Data Due to Cloud Cover: For various reasons, some attributed classes may be missing for some pixels and some time instants. In our context, the most common reason is the presence of clouds that prevented the classification for these pixels at these time instants. HMM's offer an elegant way of solving this issue by including an extra "cloud" class c from which the transition is forced out at the next time instant (i.e., by setting $p(c_t|c_{t-1}) = 0$) and for which transition to that class is forbidden as well (i.e., by setting $p(c_t|i_{t-1}) = 0$). In practice, assuming that the "cloud" class is the last one in the list, this can be done by augmenting the $\mathbf{T}$ matrix with an extra line and an extra column, with

$$\mathbf{T} = \begin{pmatrix} p(1_t|1_{t-1}) & \cdots & p(1_t|m_{t-1}) & \frac{1}{m} \\ p(2_t|1_{t-1}) & \cdots & p(2_t|m_{t-1}) & \frac{1}{m} \\ \vdots & \vdots & \vdots & \vdots \\ p(m_t|1_{t-1}) & \cdots & p(m_t|m_{t-1}) & \frac{1}{m} \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{6}$$

where the values $(1/m)$ are chosen so that each class in $\{1, \ldots, m\}$ is *a priori* equally likely to occur at the next time instant, and thus none of them is favored when including $-\ln p(i_t|c_{t-1})$ in (3). Setting null transition probabilities $p(c_t|i_{t-1})$ and $p(c_t|c_{t-1})$ will ensure that the corresponding $-\ln p(\cdot) = \infty$ in (3), so that the shortest path cannot include the cloud class after the first time instant t. Similarly, to ensure that the cloud class is not included in the Viterbi path at the initial time instant t, the $\mathbf{p}$ vector must be augmented at its end with $p(c_t) = 0$.

As there are now $m+1$ possible classes for the observations, the $\mathbf{E}$ matrix must be modified and augmented too with an extra line and column, with

$$\mathbf{E} = \begin{pmatrix} (1-\alpha)p\left(\widehat{1}_t|1_t\right) & \cdots & (1-\alpha)p\left(\widehat{1}_t|m_t\right) & \frac{1}{m} \\ (1-\alpha)p\left(\widehat{2}_t|1_t\right) & \cdots & (1-\alpha)p\left(\widehat{2}_t|m_t\right) & \frac{1}{m} \\ \vdots & \vdots & \vdots & \vdots \\ (1-\alpha)p\left(\widehat{m}_t|1_t\right) & \cdots & (1-\alpha)p\left(\widehat{m}_t|m_t\right) & \frac{1}{m} \\ \alpha & \cdots & \alpha & \frac{1}{m} \end{pmatrix} \tag{7}$$

where $\alpha = p(\widehat{c})$ is the probability of observing a cloudy pixel (that does not depend on the true class), so that $(1-\alpha)p(\widehat{i}_t|i_t)$ are the corresponding re-normalized probabilities, and where the values $(1/m)$ are chosen so that all labeled classes are again *a priori* equally likely to occur. In practice, α can

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

directly be estimated from the attributed classes using the frequency estimate

$$f(c) = \frac{1}{n \cdot (q+1)} \sum_{j=0}^{q} n\left(c_{t+j}\right) \tag{8}$$

where $n(c_{t+j})$ is the count of cloudy pixels over the set of n pixels at time instant $t + j$.

If these augmented **p**, **T** and **E** are intended to be used in the Viterbi training so that the cloud class is processed along with the other classes, one needs to reset at each iteration all probabilities associated with the cloud class in $\mathbf{p}^{(\ell)}$, $\mathbf{T}^{(\ell)}$ and $\mathbf{E}^{(\ell)}$ to the values given above. However, as long as the number of sequences affected by the presence of clouds is limited so that they do not impact much the estimates in (4), a much more convenient and faster approach is to discard at the first instance these sequences to obtain the Viterbi paths and the corresponding HMM parameters based only on the cloud-free sequences, as previously described. When this is done, the augmented version of these matrices can be used afterward for processing all cloudy sequences in a single step by accounting for the extra cloud class, without any need to estimate the parameters of the HMM again.

6) Restricting True Classes to Classified Ones: For a given sequence, any possible choice for the hidden classes is *a priori* possible at any time, even if some of these hidden classes have not been observed in the labeled sequence. In general, we are typically in situations where **T** is (strongly) diagonally dominant so that long segments of identical hidden classes are expected inside the whole sequence. Moreover, as long as the classification is reliable enough, it is also expected that **E** is diagonally dominant. Combining these two results, if we consider a segment of length S for which the true class is the same, from the geometric distribution, the mean length $E[S]$ of such a segment is given by $E[S] = 1/(1 - p(i_t|i_{t-1}))$. In contrast, from the binomial distribution the probability p^* that this true class is not part of the labeled sequence is given by $p^* = (1-p(\widehat{i_t}|i_t))^{E[S]}$. As an illustration, with values as low as $p(i_t|i_{t-1}) = 0.9$ and $p(\widehat{i_t}|i_t) = 0.5$, this leads to $E[S] = 10$ and $p^* = 0.00098$. Thus, it is unlikely that the true classes are not a subset of the observed ones in the sequence. Restricting the choice of the true classes to the subset of observed ones will ease the Viterbi training, as the computing time of the algorithm is $\mathcal{O}(m^2 \cdot q)$ so that reducing the number m of classes to be accounted for can strongly improve the performance. In practice, this consists in considering only the corresponding subset of lines and columns in the initial **T** and **E** matrices, along with the appropriate normalization so that both **T** and **E** still sum up to one per column.

Similarly, when missing data due to cloud cover are considered, only the corresponding subset of lines and columns for the augmented matrices in (6) and (7) need to be considered. Restricting the choice to the subset of labeled classes is particularly relevant when the proportion of cloud cover is important in the sequence. Ignoring this may result in choosing a Viterbi path with a single hidden class in it, different from any of the labeled classes in the sequence, just because the Viterbi path of this hidden sequence is shorter than other possible choices.

C. Validation

The China site was selected for validation as it shows intermediary land cover change rates, availability of valid and cloud-free reference observations and land cover class diversity. These characteristics allowed us to build a robust reference database by land cover interpretation. While the commission rate is easily quantified by examining areas identified as changing, it is increasingly complex to estimate the omission rate among large unchanged areas [22]. According to standards of land cover change assessment [23], the sampling should be stratified in space to target areas experiencing high rates of land cover change [24] specifically. Here, we extend these validation standards for land cover change to spurious land cover changes, i.e., discrepancies between the ML and HMM map series.

The stratified random sampling approach included three spatial strata derived from the map series themselves. The first stratum, called the agreement stratum, includes areas where both map series yield similar land cover labels for all years. The second stratum records areas with a least one cloudy year in the ML maps and the corresponding years. Finally, remaining areas and associated years with discrepant land cover labels between the ML and HMM maps constitute the third stratum of disagreement. Recorded years in the cloudy and disagreement strata correspond to the first year when the condition defining the stratum is met. A threefold assessment was thus performed to assess the ability of the HMM approach: 1) to accurately maintain a land cover label in areas that are stable over time; 2) to gap-fill clouds; and 3) to remove the spurious land cover change.

The number of samples was determined from a binomial distribution [25] with a targeted overall accuracy (OA) p of 0.8, an error margin of 0.05 and a confidence interval of 95% ($z_{1-\alpha/2} = 1.96$). A total of 245 samples allowed us to reach a trade-off between acceptable precision for the estimate and time resources at hand. The three strata were randomly allocated 49, 98 and 98 sample units, respectively. The land cover interpretation was performed with the Collect Earth Online (CEO) web interface [26] over footprints of 300 m × 300 m, which is the minimum mapping unit of the land cover maps. One land cover category was attributed per sample for each year between 2003 and 2019. Landsat and Sentinel-2 Normalized Difference Vegetation Index profiles extracted from Google Earth Engine through CEO were scanned for land cover change. The change was defined as a transition from one land cover category to another with a surface covering more than 50% of the sample footprint (i.e., 4.5 ha). For each sample, the interpretation quality was recorded with three levels of confidence (i.e., "certain," "reasonable," or "doubtful"). It is worth emphasizing that this reference database was built independently of any information about the ML and HMM map series and the sample stratum membership.

The HMM map series attributes a land cover label per pixel at each time step. In contrast, the ML time series includes

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

pixels flagged as clouds which would be systematically considered errors when validating the second stratum. For ML time series, cloud pixels were thus gap-filled with the closest valid land cover label in time (hereafter denoted naive gap-filling) to avoid artificially penalizing the ML method. The reference validation database was used afterward to validate both map series. Confusion matrices and OAs were derived for each stratum separately. F-scores were derived to highlight specific per-class accuracies– F-score being defined as $2 \times ((\text{UA} \times \text{PA})/(\text{UA} + \text{PA}))$ where UA and PA are the user and producer accuracies [27, eqs. (3) and (4)]. Additionally, weighted accuracy matrices and OAs were reported [27, eqs. (12) and (13)] to account for different pixel sampling probabilities per stratum, based on the stratum membership of each sample and the proportion of each stratum.

III. Results and Discussion

Based on the theoretical formulation and methodology proposed in Section II-B, we will present hereafter the results obtained for the three selected areas of 2520 × 2520 pixels at 300 m resolution described in Section II-A. The results will be presented jointly for the three areas (that will simply be denoted Brazil, China, and Mali hereafter). All computations were done using MATLAB R2018b on a Windows 10 workstation with 64 bits Intel Core i7 670 CPU and four cores running at 4 GHz. All computing times that are mentioned hereafter are referring to this hardware and software setup.

A. Number of Transitions

For the three areas, missing data due to the presence of clouds (labeled as such from the preprocessing cloud mask) is of varying importance. Among the 2520 × 2520 pixels, only 2.8% of them are cloud-free sequences for Brazil, while these values are equal to 30.6% and 58.9% for China and Mali, respectively. Accounting only for pixels where no clouds occurred in the sequence, one can compute the frequency distribution of these observed changes, given in Table I (first lines). One can see that the percentage of pixels for which at least one change occurred ranges from 39.6% to 51.4%. Multiple changes (number of transitions above one) over the sequences are frequent, with values equal to 44.3%, 36.0%, and 43.4% for Brazil, China, and Mali, respectively. These values grossly overestimate the real changes that occurred, as most of these observed changes are expected to be caused by classification errors that vary from 1 year to another. This high number of reported transitions for the three areas cannot be considered realistic over a 17-year time span. Fig. 2 (left) presents this number of changes for pixels with less than two years of clouds in the sequence.

Looking now at the number of transitions computed from the Viterbi paths (i.e. the most likely sequence of classes for each pixel) as given in Table I (second lines) and Fig. 2 (right), the results look much more in agreement with what one would expect, where about 80% of the pixels correspond to unchanged pixels over the 17-year time span. Multiple changes are now less likely and correspond to 3.9%, 4.2%, and 6.7% of the pixels for Brazil, China, and Mali, respectively. These findings are also in close agreement with those [13] obtained in the southwestern United States, where they observed a dramatically reduced number of transitions when HMM is used to account for classification errors. From an assessment based on a database of nearly 2000 sites distributed worldwide that

TABLE I

Frequencies (in %) for the Number of Transitions (#) Over the Years 200–2019 at the Three Study Sites. First Lines (Obs) Correspond to the Attributed Classes, While Second Lines (HMM) Correspond to the Viterbi Paths

Brazil

#	0	1	2	3	4	5	6	7	8+
Obs	48.6	7.1	11.7	8.7	7.1	6.0	4.2	3.1	3.5
HMM	83.5	12.6	3.3	0.6	0.05	0	0	0	0

China

#	0	1	2	3	4	5	6	7	8+
Obs	60.4	3.6	13.8	4.6	6.3	3.6	3.2	1.9	2.6
HMM	82.1	13.7	3.9	0.3	0.02	0	0	0	0

Mali

#	0	1	2	3	4	5	6	7	8+
Obs	54.2	2.4	15.5	4.5	8.1	4.3	4.6	2.7	3.7
HMM	79.9	13.4	6.0	0.6	0.05	0	0	0	0

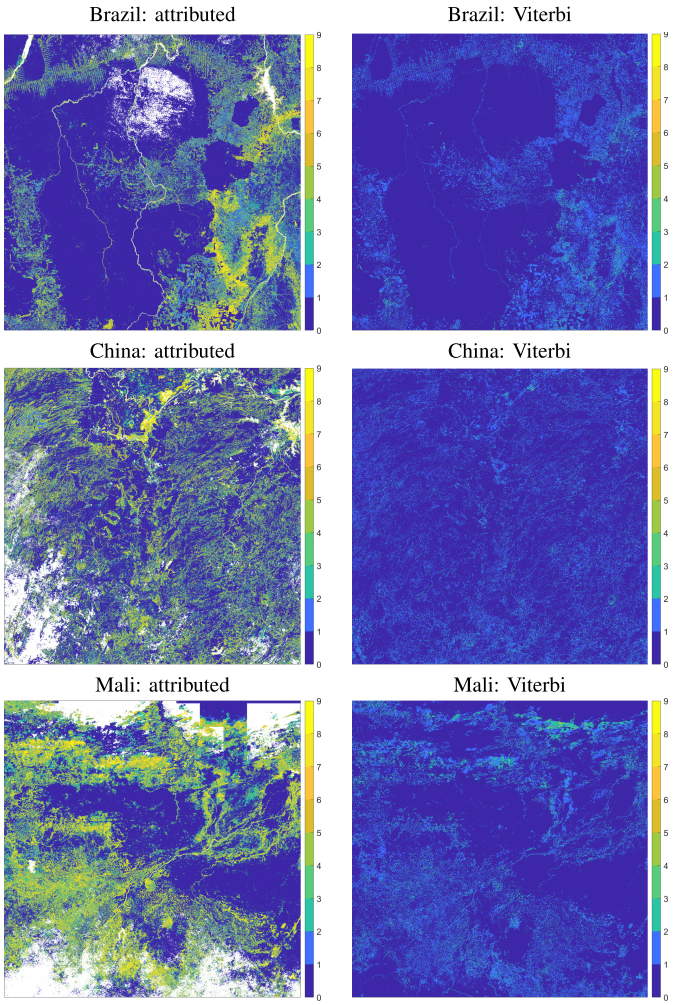


Fig. 2. Number of class transitions over the years 2003–2019 at the three study sites. Left-side figures are obtained from the attributed classes, while right-side figures are obtained from the Viterbi paths. White pixels correspond to pixels with clouds in the sequence for at least 2 years. Sharp borders relate to the tiles used to build the images.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

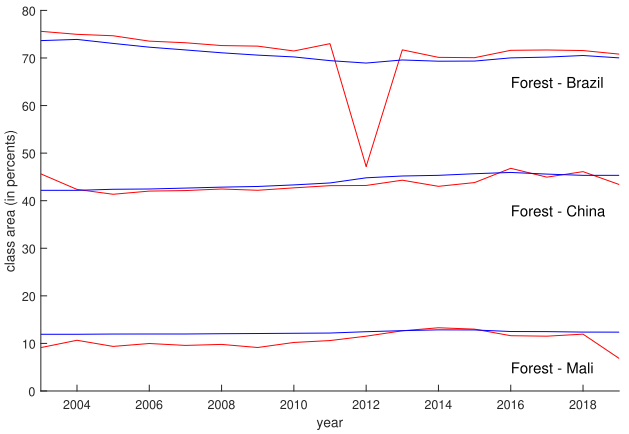


Fig. 3. Evolution of the area (in percents) covered by the Forest class at Brazil, China and Mali sites over the 2003–2019 time span. Red lines refer to the estimation using the raw data, while blue lines refer to the estimation from the HMM results.

were selected to be homogeneous and stable through time, they concluded that HMM properly corrected these errors. Besides reducing the number of transitions, HMM also improves the temporal consistency of the classes, as illustrated for the Forest class in Fig. 3. Using the initial data (with missing values due to clouds), it can be seen that fluctuations of the area covered by the class are randomly varying from one year to another. On the opposite, the HMM results yield smooth variations over time, as expected. The impact on the cloud cover can be dramatic for Brazil, where only 2.8% of the area was visible in 2012, thus leading to biased estimates of the class area at that date, as the visible part is not representative of the whole area for that class. For Mali, the Forest class percentage estimated from HMM results is almost constant over 2003–2019. The range of variations is close to 5% over the same period using the initial classes.

The final output of the HMM model is the sequence of maps that maximize the likelihood of the labels over these 17 years. An appropriate visualization of this output would require presenting all maps simultaneously, either side by side or inside a video sequence. This is, however, cumbersome and does not ease a visual synthesis of the results. One solution is to rely on the use of (3) where the Viterbi path $\mathbf{i}^{(\infty)}$ can be plugged. Doing this for all pixels allows us to present a single map of the $-\ln p(\mathbf{i}^{(\infty)}|\widehat{\mathbf{i}})$ values for the whole sequence over each area, as presented in Fig. 3. These maps incorporate at the same time the information about the number of transitions and their corresponding probabilities, as measured from the various $-\ln p(i_{t+j}^{(\infty)}|i_{t+j-1}^{(\infty)})$'s along the Viterbi path, along with the corresponding emission probabilities. Low values are typically associated with unchanged pixels, while high values are associated with multiple and/or unlikely changes. As convoluted as this global measure might appear, it nevertheless allows the user to identify characteristic patterns at first glance, e.g., deforestation patterns and modifications of lake extent over the years, and spot places where changes took place easily.

B. Emission and Transition Probabilities

Using only the sequences of labeled maps in a Viterbi training algorithm by discarding at a first instance those that

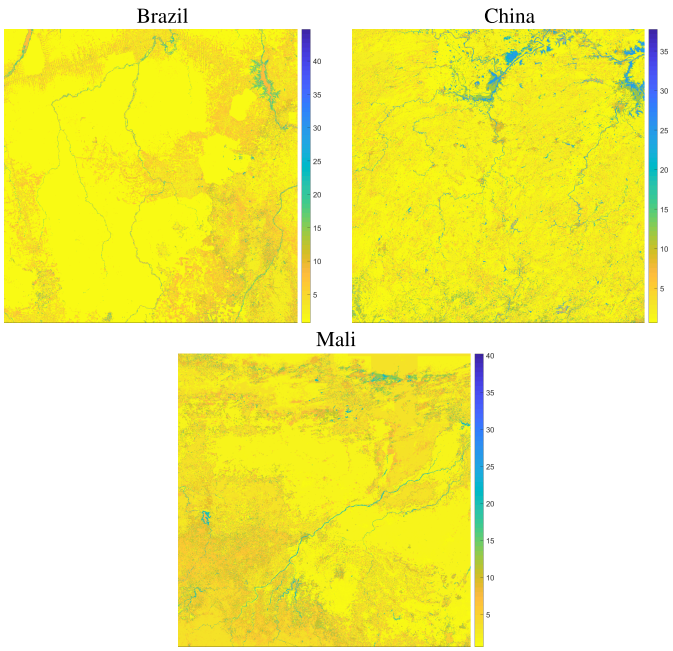


Fig. 4. Maps of the $-\ln p(\mathbf{i}^{(\infty)}|\widehat{\mathbf{i}})$ values over the three study sites, as computed from the estimated Viterbi paths $\mathbf{i}^{(\infty)}$ and probability estimates $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$.

are not cloud-free, one can jointly obtain the estimates $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ of the HMM parameters, along with the corresponding Viterbi paths $\mathbf{i}^{(\infty)}$ that were already presented and discussed above. We will consider initial guesses $\mathbf{p}^{(0)}$, $\mathbf{T}^{(0)}$ and $\mathbf{E}^{(0)}$ that are given by

$$\begin{aligned}\mathbf{p}^{(0)} &= \frac{1}{m}\mathbf{1}\\ \mathbf{T}^{(0)} &= \alpha_T \cdot \mathbf{I} + \frac{1-\alpha_T}{m}\mathbf{1}\mathbf{1}'\\ \mathbf{E}^{(0)} &= \alpha_E \cdot \mathbf{I} + \frac{1-\alpha_E}{m}\mathbf{1}\mathbf{1}'\end{aligned} \tag{9}$$

with $\mathbf{1}$ the unit 6×1 vector and $\mathbf{1}\mathbf{1}'$ the unit 6×6 matrix, where $m = 6$ for our study. Choosing $\alpha_T \geq 0.5$ and $\alpha_E \geq 0.5$ will lead to diagonally dominant matrix for $\mathbf{T}^{(0)}$ and $\mathbf{E}^{(0)}$. The impact of the specific choices for α_T and α_E will be investigated later on. Starting close enough to the expected values helps to reduce the number of iterations for the Viterbi training and limits the risk of obtaining poor estimates. We choose to set $\alpha_T = 0.98$ and $\alpha_E = 0.8$ as initial guesses, that can be considered as realistic values. Indeed, change rates are typically in the range of a few percent per year (even if their values depend on the land cover classes). For any of the $m = 6$ class, it corresponds to a total transition rate equal to 1.67% toward any other class and a producer's accuracy equal to 83.3%. For the sake of simplicity, the same initial guesses will be considered for the three areas. It is worth emphasizing that these values are only initial guesses that do not presume about the final estimates. Based on these initial guesses, the number of iterations needed to obtain the final estimates $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$, and $\mathbf{E}^{(\infty)}$ (i.e. the final values so that no more changes occurred in the Viterbi paths) where equal to 28, 41, and 20 for Brazil, China, and Mali, respectively, where each Viterbi training iteration requires a computing time that

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

TABLE II

RESULTS FOR THE BRAZIL STUDY SITE, SHOWING THE VITERBI TRAINING PROBABILITY ESTIMATES (IN %) $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$, AND $\mathbf{E}^{(\infty)}$, ALONG WITH THE AVERAGE $\overline{\mathbf{p}}$, $\overline{\mathbf{T}}$, AND $\overline{\mathbf{E}}$ OF THE VITERBI TRAINING ESTIMATES FOR 100 SIMULATIONS THAT ARE USING $\mathbf{p}^{(\infty)}$, $\mathbf{E}^{(\infty)}$, AND $\mathbf{T}^{(\infty)}$ AS EMISSION AND TRANSITION PROBABILITIES. LABELS ARE C = CROPLAND, F = FOREST, G = GRASSLAND, W = WATER, S = SETTLEMENT, AND O = OTHER

$\mathbf{E}^{(\infty)}$

	C	F	G	W	S	O
C	86.09	0.55	2.68	2.17	10.73	0.86
F	8.49	98.90	2.14	7.23	0.23	1.73
G	4.27	0.35	91.03	2.46	19.55	6.12
W	0	0.01	0	58.82	0	0.02
S	0.08	0	0.05	0	48.10	0.16
O	1.08	0.20	4.09	29.34	21.38	91.11

$\overline{\mathbf{E}}$

	C	F	G	W	S	O
C	86.14	0.53	2.72	2.24	10.48	0.88
F	8.33	98.94	2.19	7.26	0.17	1.75
G	4.35	0.33	90.95	2.30	19.81	6.09
W	0	0.01	0	60.29	0	0.02
S	0.08	0	0.05	0	51.32	0.18
O	1.09	0.19	4.08	27.91	18.23	91.08

$\mathbf{T}^{(\infty)}$

	C	F	G	W	S	O
C	97.70	1.10	1.50	0.70	5.20	0.40
F	1.50	98.60	1.80	0.60	0	1.00
G	0.60	0.30	96.10	0	0.10	0.80
W	0	0	0	94.10	0	0
S	0	0	0	0.10	89.50	0.10
O	0.2000	0.10	0.60	4.50	5.10	97.70

$\overline{\mathbf{T}}$

	C	F	G	W	S	O
C	98.20	1.00	1.30	0.60	4.50	0.30
F	1.30	98.70	1.80	0.50	0	1.00
G	0.50	0.20	96.50	0	0	0.70
W	0	0	0	95.50	0	0
S	0	0	0	0.10	92.40	0.10
O	0.1000	0.10	0.50	3.30	3.10	98.00

$\mathbf{p}^{(\infty)}$

C	F	G	W	S	O
15.43	48.06	24.15	0.05	0.16	12.15

$\overline{\mathbf{p}}$

C	F	G	W	S	O
15.53	50.19	22.32	0.04	0.15	11.78

directly depends on the number of pixels to be processed and the number of distinct labeled classes in each pixel sequence. On average, this computing time is about 1.2 10^{-4} second per pixel (for 17 years' sequences in our study). The corresponding results are given in Tables II–IV. The detailed assessment of these results is delayed to Section III-C for easing the presentation. Looking first at the emission probabilities $\mathbf{E}^{(\infty)}$, most land cover types are associated with low classification errors (producer's accuracy above 80%) except for Water and Settlement in Brazil, Grassland and Water in China, and Water in Mali. Classes with low estimated accuracy (below 60%) typically have low estimated frequencies (below 0.4%). It is clear for very marginal classes that only a few mislabeled pixels often due to omission errors significantly impact class accuracy.

TABLE III

RESULTS FOR THE CHINA STUDY SITE. SAME LEGEND AS FOR TABLE II

$\mathbf{E}^{(\infty)}$

	C	F	G	W	S	O
C	96.87	4.74	22.04	20.69	13.19	17.68
F	2.67	95.05	25.99	9.29	0.16	9.68
G	0.04	0.02	42.77	9.38	1.06	0.58
W	0.02	0.01	1.61	34.81	0.29	0.69
S	0.24	0.01	3.89	11.70	84.56	1.29
O	0.16	0.16	3.70	14.13	0.73	70.07

$\overline{\mathbf{E}}$

	C	F	G	W	S	O
C	96.78	4.77	20.62	21.32	11.13	16.99
F	2.71	95.01	26.57	9.45	0.16	9.28
G	0.04	0.02	43.35	9.18	1.12	0.62
W	0.02	0.01	1.54	34.26	0.28	0.73
S	0.26	0.01	4.05	11.44	86.54	1.38
O	0.18	0.18	3.86	14.35	0.77	71.00

$\mathbf{T}^{(\infty)}$

	C	F	G	W	S	O
C	98.53	0.79	2.35	1.13	4.33	1.22
F	1.05	99.02	0.83	1.29	0.07	1.70
G	0.04	0	95.34	0.09	0.35	0.02
W	0	0	0.02	95.53	0.01	0.28
S	0.21	0	1.24	1.54	95.09	0.27
O	0.16	0.18	0.23	0.41	0.14	96.51

$\overline{\mathbf{T}}$

	C	F	G	W	S	O
C	98.75	0.67	1.50	0.81	3.53	0.83
F	0.89	99.17	0.37	1.07	0.05	1.32
G	0.03	0	97.16	0	0.23	0.01
W	0	0	0	96.94	0	0.11
S	0.19	0	0.87	1.06	96.07	0.20
O	0.13	0.15	0.10	0.12	0.11	97.53

$\mathbf{p}^{(\infty)}$

C	F	G	W	S	O
53.31	43.27	0.39	0.32	1.18	1.54

$\overline{\mathbf{p}}$

C	F	G	W	S	O
53.40	43.00	0.37	0.33	1.19	1.70

Looking now at the transition probabilities $\mathbf{T}^{(\infty)}$, all diagonal values (that corresponds to no change between two successive years) are above 95% for almost all classes and overall areas. These results agree with the general observation that annual rates of changes are typically low when estimated over large areas. For example, for the Brazilian federal state of Para that includes our study site [28], the mean annual deforestation rate using 30-m Landsat long-term time series for the period spanning from 1984 to 2014 was estimated to be 0.66%.

C. Viterbi Training Assessment

Clearly, all probabilities in $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ have been directly estimated from the labeled maps only, without accounting for any strong prior information on the classification performance

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

RESULTS FOR THE MALI STUDY SITE. SAME LEGEND AS FOR TABLE II

$\mathbf{E}^{(\infty)}$

	C	F	G	W	S	O
C	96.72	2.06	5.96	20.48	11.75	5.54
F	0.14	85.76	0.07	8.93	1.15	1.64
G	1.71	0.02	91.46	3.33	0.61	1.48
W	0.02	0.03	0.01	45.22	0.15	0.04
S	0.02	0	0.01	1.18	85.82	0.01
O	1.38	12.12	2.49	20.86	0.52	91.30

$\overline{\mathbf{E}}$

	C	F	G	W	S	O
C	96.62	2.05	5.94	20.09	10.88	5.43
F	0.15	86.21	0.07	8.85	1.19	1.71
G	1.77	0.02	91.38	3.38	0.56	1.54
W	0.02	0.03	0.01	45.16	0.12	0.04
S	0.03	0	0.01	1.22	86.73	0.01
0	1.41	11.69	2.58	21.31	0.53	91.27

$\mathbf{T}^{(\infty)}$

	C	F	G	W	S	O
C	99.20	0.36	1.61	1.47	2.29	1.48
F	0.05	97.70	0.01	0.57	0.24	0.61
G	0.42	0.01	97.42	0.36	0.25	0.71
W	0.02	0.11	0.01	96.89	0.70	0.02
S	0.01	0	0	0.26	96.43	0
O	0.29	1.82	0.95	0.44	0.09	97.17

$\overline{\mathbf{T}}$

	C	F	G	W	S	O
C	99.33	0.32	1.36	0.95	1.93	1.28
F	0.05	98.16	0.01	0.36	0.24	0.51
G	0.35	0	97.86	0.28	0.19	0.57
W	0.02	0.09	0.01	98.04	0.55	0.01
S	0.01	0	0	0.20	97.01	0
O	0.25	1.43	0.77	0.17	0.07	97.63

$\mathbf{p}^{(\infty)}$

C	F	G	W	S	O
57.19	4.55	19.22	0.36	0.14	18.53

$\overline{\mathbf{p}}$

C	F	G	W	S	O
58.18	4.54	18.41	0.36	0.15	18.37

or the expected rates of change over the area. Although the initial choice for $\mathbf{T}^{(0)}$ and $\mathbf{E}^{(0)}$ are consistent with those generally found in the literature, it is known too that Viterbi training estimates are biased in general and can be sensitive to the initial guesses. It is thus important to assess these aspects in our context. We propose to do so here using a simulation approach. Let us consider that $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$, and $\mathbf{E}^{(\infty)}$ (see left of Tables II–IV) are the correct theoretical probabilities. From $\mathbf{T}^{(\infty)}$, one can thus first simulate a fictitious sequence $\mathbf{i}$ of true classes, from which in turn a fictitious sequence $\widehat{\mathbf{i}}$ of labeled classes can be simulated using $\mathbf{E}^{(\infty)}$. Simulating this way and independently a large set of sequences allows us to use a Viterbi training on this set for estimating the emission and transition probabilities. Repeating this simulation procedure a large number of times makes it possible to compute the bias and the variance of the corresponding Viterbi training estimates that need to be compared to the reference values $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ that were used for the simulations, of course. Clearly, the $\mathbf{p}^{(\infty)}$, $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ used in the simulation do not necessarily correspond to the correct values for $\mathbf{p}$, $\mathbf{T}$ and $\mathbf{E}$, but as long as the Viterbi training is assumed

TABLE V

SUMMARY OF THE ESTIMATED BIAS b, STANDARD DEVIATION S AND RMSE $= \sqrt{b^2 + S^2}$ FOR THE VITERBI TRAINING ESTIMATES OF THE EMISSION $\mathbf{E}^{(\infty)}$ AND TRANSITION $\mathbf{T}^{(\infty)}$ PROBABILITIES (IN %) AT THE THREE STUDY SITES. ONLY AVERAGE VALUES (LEFT PART) AND MAXIMUM VALUES (RIGHT PART) ARE GIVEN FOR THE SAKE OF CONCISENESS

$\mathbf{E}^{(\infty)}$

	$\overline{b}$	$\overline{S}$	$\overline{\text{RMSE}}$	$\max(b)$	$\max(S)$	$\max(\text{RMSE})$
Brazil	0.30	0.19	0.39	3.22	1.32	3.32
China	0.32	0.14	0.37	2.06	0.51	2.07
Mali	0.13	0.10	0.17	0.91	0.50	0.98

$\mathbf{T}^{(\infty)}$

	$\overline{b}$	$\overline{S}$	$\overline{\text{RMSE}}$	$\max(b)$	$\max(S)$	$\max(\text{RMSE})$
Brazil	0.30	0.09	0.33	2.88	0.61	2.92
China	0.31	0.04	0.32	1.83	0.19	1.83
Mali	0.18	0.04	0.19	1.15	0.22	1.16

to provide meaningful results, the corresponding simulated sequences of land covers can be considered as realistic hypothetical real sequences for each of the three areas. The corresponding results are given in the right of Tables II–IV, where a set of 250 000 sequences (of 17 years length each) was simulated 100 times for the three areas. For each simulated set, a Viterbi training is conducted for obtaining the $\mathbf{p}$, $\mathbf{T}$, and $\mathbf{E}$ estimates. From the 100 repetitions, the average estimates $\overline{\mathbf{p}}$, $\overline{\mathbf{T}}$, and $\overline{\mathbf{E}}$ are obtained. The differences between the values on the left and right parts of the table are the (estimated) biases for the probabilities. The standard deviation of these estimates can be computed along with the corresponding root mean square errors (RMSE). For each study site, the maximum and mean values of the bias, standard deviation, and RMSE are given in Table V.

As seen from Table V, the Viterbi training estimates are indeed biased in most instances (as confirmed from Student's t-tests; results not shown here), but on average, these biases are close to or lower than 0.3% both for the emission and transition probabilities and for all areas. It can also be seen that the main contribution to the total error (measured here with the RMSE) is precisely due to the bias, as standard deviations are comparatively low in all instances. The maximum values for the bias in Table V can be large (up to 3%) in some instances, but these high values are typically concentrated over rare classes. For Brazil, all biases exceeding 1% for the emission (4 out of the 36) and transition (4 out of the 36) probabilities are observed for the Water and Settlement. For China, the same result holds, where all bias exceeding 1% for the emission (3 out of the 36) probabilities and transition (3 out of the 36) probabilities are observed for Grassland, Water, and Settlement. For Mali, a single bias exceeding 1% is observed for the transition probability of Water. The fact that high biases are concentrated over low-frequency classes is easily seen from Fig. 4, which presents the bias for the emission and transition probabilities as a function of the class frequency. Typically, as soon as a class frequency exceeds 1.5%, the corresponding biases do not exceed 0.5%. In parallel to this, the asymptotic properties of these estimates were investigated by varying the number of sequences used for the simulation. For the sake of brevity, corresponding results are

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

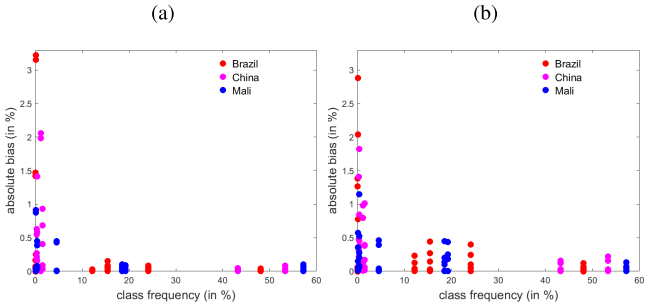


Fig. 5. Values of the absolute bias for the probability estimates as a function of the class frequency, as computed from 100 simulated sets of 250 000 sequences. (a) Emission probabilities $\overline{\mathbf{E}}$. (b) Transition probabilities $\overline{\mathbf{T}}$.

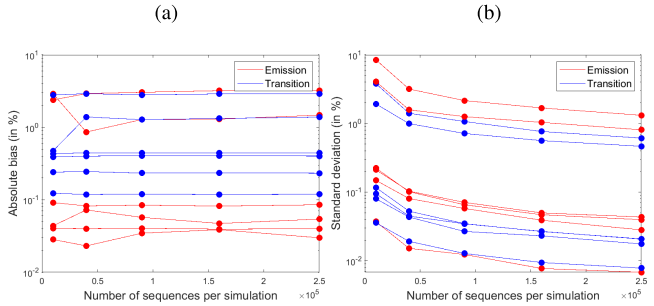


Fig. 6. Values of (a) absolute bias and (b) standard deviation for the diagonal probability estimates in $\overline{\mathbf{E}}$ and $\overline{\mathbf{T}}$ as a function of the number of sequences in each simulation (for the Brazil area only), as computed from 100 simulated sets of such sequences. Vertical scales are logarithmic for the sake of readability.

given in Fig. 5 for Brazil only, by focusing on the diagonal elements of the $\overline{\mathbf{T}}$ and $\overline{\mathbf{E}}$ matrices (results are qualitatively similar for the other probabilities and the other areas). Without surprise, the standard deviation of the probability estimates decreases with the number of sequences. More surprisingly, the bias is constant regardless of the number of sequences (Fig. 6), so these probability estimates cannot be considered asymptotically unbiased regarding the number of sequences used to estimate them. The same conclusion is true if the length of the simulated sequences is increased (results not shown here), so that biased estimates are unavoidable here.

As previously mentioned, another potential issue is the dependency that these results may have on the choice for the initial guesses $\mathbf{T}^{(0)}$ and $\mathbf{E}^{(0)}$.

To investigate this aspect, values for α_T and α_E were varied over the $[0.9, 0.99]$ and $[0.6, 0.95]$ intervals, respectively, and the corresponding estimates of $\mathbf{T}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ were computed. The maximum difference between probability estimates is typically below 0.01%, so they can be considered negligible by comparison with the RMSE. It is indeed true that the initial guesses can impact the result when the number of sequences is limited, especially for low-frequency classes (results not shown here). Still, this impact is limited by comparison with the other sources of errors and vanishes as the number of sequences increases. It is worth mentioning that considering a $[0.9, 0.99]$ interval for α_T in this simulation study might seem restrictive, as any value in the $[0, 1]$ interval is theoretically possible. However, using such a large interval would not make much sense as: 1) it does not agree with the typically observed change rates; 2) it would dramatically slow down the convergence of the algorithm if one starts from an initial guess that is clearly far away from the expected final value; and 3) it increases the risk of convergence issues without necessity.

As long as a class is reasonably well represented on the map and that enough sequences are considered in the Viterbi training, tentative conclusions are that biases remain quite limited both for the emission and transition probabilities estimates, with low variances and dependence on initial guesses that can be neglected as long as realistic initial guesses are selected. If one considers that the hypotheses at the root of the HMM are fulfilled, these results emphasize that automatic learning of the HMM parameters using a Viterbi training is a quite reliable methodology in our context, thus alleviating the burden of using the much more demanding Baum-Welch algorithm. Of course, these results do not prove that HMM is a relevant modeling approach in our context, as this would require a careful assessment of all the HMM hypotheses, along with an extensive (and costly) external validation procedure, that could not depend on the use of the HMM model itself. However, as mentioned and shown in [13], previous uses of an HMM approach already lead to convincing results about its benefit by comparison with other post-classification stabilization heuristics. Moreover, the hypotheses behind an HMM are not very stringent and could be alleviated if needed, as discussed in Section III. While most existing land cover change assessment approaches either focus specifically on the change using multi-year classification methods (e.g. [14], [15], [28] or rely on a validation strategy to adjust the derived change area estimate [23], the proposed HMM framework allows us to take advantage of the mapping exercise completed independently at regular time intervals.

TABLE VI

DISTRIBUTION OF THE REFERENCE SAMPLES PER LAND COVER CATEGORIES AND STRATUM

	C	F	G	W	S	O	TOTAL
Agreement	22	26	0	0	0	1	49
Cloudy	27	56	3	3	5	4	98
Disagreement	19	67	1	4	6	1	98
TOTAL	68	149	4	7	11	6	245

D. HMM Performance

All samples were interpreted with a high level of confidence, tagged as "certain" so that all samples could be used in the performance evaluation. The China landscape is mainly composed of forest land and cropland, as shown in the reference database (Table VI). Among remaining classes, the settlement category slightly dominates.

Strata of agreement, cloudy areas, and disagreements are distributed in proportions of 19%, 46%, and 35%, respectively [Fig. 7(a)]. All years are represented at least once in the temporal stratification [Fig. 7(b)]. The map series assessment is done for the specific years recorded in the stratification. As all land cover labels are similar in time for both map series in the stable stratum, 2019 was arbitrarily chosen for the assessment.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

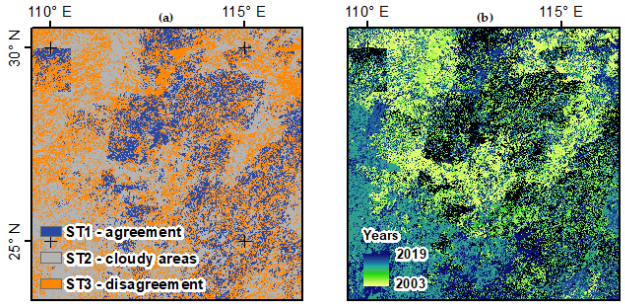


Fig. 7. Stratification. (a) Distribution of the 3 strata aiming at assessing the performance of the HMM method over stable areas, as a gap-filler in cloudy areas and as corrections for spurious changes. (b) Years in the 2003–2019 period associated with the pixels from the 2nd and 3rd strata. Stable areas are illustrated in black.

Confusion matrices and OAs are first reported per stratum to better highlight differences in performance between methods according to the context (Table VII). The ML and HMM map series yield identical high OA of 83.67% in the agreement stratum when no change is observed over time. Not surprisingly, the HMM correctly maintains land cover labels in stable areas. In cloudy areas, the OA of both map series drops to similar values of about 70%. The HMM approach can thus be considered as an efficient method for gap-filling categorical map series. Over the remaining areas (i.e. the disagreement stratum), the OA almost doubled from 31% to 60% with the HMM approach. The major misclassifications in the ML map series occurred between the cropland and forest categories. In this area of China, croplands are often cultivated in narrow valleys surrounded by forests. These mixed pixels switch erroneously from one class to the other from year to year, leading to temporally inconsistent classifications. Taking the temporal context into account with the HMM approach, dramatic improvements in the F-score of 19% and 31% were obtained for cropland and forest classification accuracy, respectively. Overall, the HMM approach increases the stability of yearly land cover map series processed independently by reclassifying temporally unstable cropland pixels as forest land. This increase in consistency, also reported in [13], allows reducing spurious changes in the classification series. Although the HMM performance should be interpreted with considerable caution for the weakly represented categories in the reference sample (Table VI), increases of 40% and 58% of F-score were also observed for water and settlement, respectively.

To account for the uneven pixel sampling probabilities per stratum, we used strata weights equivalent to the respective stratum proportions and obtained weighted accuracy matrices. The derived OAs (Table VIII) show that HMM maps significantly surpass the naively gap-filled ML maps by 12%, which is well above the 5% margin of error. These figures evidenced that the benefit of HMM is to increase the consistency of ML land cover map series by reducing erroneous land cover changes. In China, HMM mainly reduces the misclassification between the most likely forest and cropland classes by 12%. These results encourage testing the HMM method at a larger scale under greater landscape diversity and class occurrence probabilities.

TABLE VII

CONFUSION MATRICES AND OAS DERIVED PER STRATUM FOR (a) THE ML MAP SERIES 2003–2019 AND (b) THE HMM MAP SERIES 2003–2019. UAS, PAS

Agreement stratum

	Reference							Reference					
(a)	C	F	G	W	S	O	**(b)**	C	F	G	W	S	O
C	22	7	0	0	0	1	C	22	7	0	0	0	1
F	0	19	0	0	0	0	F	0	19	0	0	0	0
G	0	0	0	0	0	0	G	0	0	0	0	0	0
W	0	0	0	0	0	0	W	0	0	0	0	0	0
S	0	0	0	0	0	0	S	0	0	0	0	0	0
O	0	0	0	0	0	0	O	0	0	0	0	0	0
OA			0.84				**OA**			0.84			

Cloudy stratum

(a)	C	F	G	W	S	O	**(b)**	C	F	G	W	S	O
C	25	15	2	2	2	4	C	25	13	2	2	2	4
F	2	40	1	0	0	0	F	2	42	1	0	0	0
G	0	0	0	0	0	0	G	0	0	0	0	0	0
W	0	0	0	1	0	0	W	0	0	0	1	1	0
S	0	0	0	0	3	0	S	0	0	0	0	2	0
O	0	1	0	0	0	0	O	0	1	0	0	0	0
OA			0.70				**OA**			0.71			

Disagreement stratum

(a)	C	F	G	W	S	O	**(b)**	C	F	G	W	S	O
C	7	40	0	0	2	0	C	11	19	1	2	2	1
F	10	22	1	1	1	1	F	6	43	0	0	0	0
G	0	0	0	1	1	0	G	0	1	0	0	0	0
W	0	1	0	0	0	0	W	0	0	0	1	0	0
S	0	0	0	2	1	0	S	0	0	0	0	4	0
O	2	4	0	0	1	0	O	2	4	0	1	0	0
OA			0.31				**OA**			0.60			
UA	0.14	0.61	0	0	0.33	0	UA	0.31	0.88	0	1	1	0
PA	0.37	0.33	0	0	0.17	0	PA	0.58	0.64	0	0.25	0.67	0
FS	0.21	0.43	0	0	0.22	0	FS	0.40	0.74	0	0.40	0.80	0

E. Applicability for Global Land Cover Mapping

Although all pixels but cloudy ones were used in the Viterbi training algorithm for our three study cases, this would not be needed in a mapping framework over much larger areas. Indeed, the Viterbi training is only required to estimate the HMM parameters, which can be used afterward for the Viterbi decoding over other areas, as long as the same set of parameters can reasonably characterize these areas. Moreover, there is no requirement of spatial adjacency for the pixels used in the Viterbi training. Any subset of pixels in the area of interest could be used for this Viterbi training stage. In other words, only a fraction of the pixels need to be used to estimate the HMM parameters. Depending on the variability and complexity of the land cover patterns, the optimal fraction of pixels to be used for the parameters estimation and the spatial extent over which the same set of parameters can be used need to be assessed to allow reliable estimations while avoiding at the same time a useless computational burden. As the optimal fraction is expected to vary by region, no absolute guidelines can be provided at this stage. However, from our study site in Brazil, where only 2.8% of the pixels were cloud-free at all dates, this fraction does not need to exceed a few percent, as long as these pixels are representative of an area for which the same parameters are supposed to

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII

UNBIASED CONFUSION MATRICES AND OAS DERIVED USING ALL SAMPLES ON (a) ML MAP SERIES 2003–2019 AND ON (b) THE HMM MAP SERIES 2003–2019. ROWS AND COLUMNS REPRESENT THE MAP SERIES AND REFERENCE CLASSIFICATION, RESPECTIVELY

Weighted confusion matrices													
	Reference							Reference					
(a)	C	F	G	W	S	O	**(b)**	C	F	G	W	S	O
C	18.1	22.2	0.9	0.9	1.6	2.0	C	19.5	14.0	1.3	1.6	1.6	2.4
F	4.4	29.7	0.8	0.4	0.4	0.4	F	3.0	38.0	0.5	0.0	0.0	0.0
G	0.0	0.0	0.0	0.4	0.4	0.0	G	0.0	0.4	0.0	0.0	0.0	0.0
W	0.0	0.4	0.0	0.5	0.0	0.0	W	0.0	0.0	0.0	0.8	0.5	0.0
S	0.0	0.0	0.0	0.7	1.7	0.0	S	0.0	0.0	0.0	0.0	2.3	0.0
O	0.7	1.9	0.0	0.0	0.4	0.0	O	0.7	1.9	0.0	0.4	0.0	0.0
OA	0.56							0.68					

apply. As previously discussed, the main limitation probably comes from the rare classes, which need to be sufficiently represented to allow reliable estimations.

It is also possible to estimate the total computing time that a 300-m resolution global mapping would require. The estimated computing time was about $1.2–10^{-4}$ second/sequence for each pixel of 0.09 km^2 area for the Viterbi decoding. Accounting for the fact that the global terrestrial area covers about 148 940 000 km^2, this corresponds to a total computing time of about 55 h on a high-end personal computer as the one used in this study. This computing time will quadratically increase with the linear spatial resolution and the number of classes. However, it should remain a competitive and tractable approach for a global mapping at this 300-m resolution using a possibly richer labeling (i.e. more than the six classes used in our study) as long as the appropriate hardware resources are at hand.

It is worth mentioning that increasing the number of classes will increase the number of possible transitions between these classes, but this does not necessarily increase the rate of change as long as one stays inside the same kind of classes. For example, including additional classes associated with specific crops inside the generic "cropland" class will obviously increase the change rates as crops may change over time at the same place, but this will not impact the transition from "cropland" to, say, "forest" or "settlement." Moreover, as we processed here data that are resulting from a preliminary classification, using more classes in the HMM algorithm would assume that these classes can reasonably be obtained from the classification procedure itself. For studies conducted at a 300-m resolution, a limited number of classes makes sense, as they can reasonably be separated during the classification and validation steps and are helpful for land use management. This is also the reason why they are classically used at that scale.

IV. CONCLUSION

We demonstrated a gain of the OA of 12% in China for the HMM approach. We also assessed the benefits of an HMM approach for processing time series of annual satellite images using a combination of real datasets and simulations. Yet, like any model, it is subject to limitations inherent to major underlying hypotheses, i.e. the probability of observing a state at a given time instant only depends on the previously observed state, with emission and transition probabilities that are time-invariant. One can, however, easily think about more complex temporal dependencies involving cyclic or long-term transitions between land use classes, the impact of land use management policies on transition probabilities (e.g., large-scale reforestation policies), and technological improvements (i.e., lower classification errors) on emission probabilities. This might justify considering more complex models like higher order and variable-order Markov chains [29], [30], or time-varying models [31]. At this stage, it is unclear how these improved models could outperform a simpler approach and if they remain computationally tractable in our context, which typically requires processing millions of chains.

We emphasized that HMM parameters can directly be estimated from the classified pixels without external information about the emission transition probabilities. Nevertheless, it is expected that accounting for expert-based rules could improve the quality of the results. Clearly, some transitions are unexpected (e.g. Water to Forest) but might still occur if the corresponding transition probabilities are not set equal to 0. Forbidding undesirable transitions can be done similar to removing clouds, i.e., forcing entries equal to 1 or 0 in the appropriate columns of the matrices. Similarly, if it turns out that transition probabilities can objectively be identified as overestimated and lead to too numerous transitions in the sequences, their values can be arbitrarily decreased or set by the user. This offers some flexibility and control over the results coming out from the HMM. The major limitation is that, as is, HMM cannot account for more complex rules that involve more than two successive time instant. If such rules need to be accounted for, this can only be done by applying these rules afterward to the HMM results.

As presented here, our HMM is also applied on a per-pixel basis so that all pixels are processed separately and independently of each other. No constraint on the spatial consistency of the results is thus accounted for. If required, improving this consistency can easily be done by sliding windows at each time instant to filter out classes that, e.g., are unlikely to occur in an isolated way. Though the HMM could be modified to process all pixels in such a spatial window jointly, this would require a modification of the algorithm along with a proper selection of the window to be used, which may heavily depend on the spatial pattern of each class in the area. Extension of the classical Markov chain time series to Markov random fields have been proposed in [32] and [33].

In general, the HMM approach needs to be considered inside a more general framework that could involve additional postprocessing steps based on other requirements. It remains, however, a valuable tool for automatically processing time series of noisy image classification outputs, as illustrated in this article.

APPENDIX
HMM DERIVATION

Using basic probability calculus, we can write that

$$p\left(\mathbf{i}|\hat{\mathbf{i}}\right) = \frac{p\left(\mathbf{i},\hat{\mathbf{i}}\right)}{p\left(\hat{\mathbf{i}}\right)} \propto p\left(\mathbf{i},\hat{\mathbf{i}}\right) = p(\mathbf{i})\,p\left(\hat{\mathbf{i}}|\mathbf{i}\right) \tag{A.1}$$

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

where $p(\widehat{\mathbf{i}})$ is a constant value than does not depend on $\mathbf{i}$. Let us first write the joint probability $p(\mathbf{i})$ as

$$p(\mathbf{i}) = p(i_t)p(i_{t+1}|i_t)p(i_{t+2}|i_t, i_{t+1}) \cdots p\left(i_{t+q}|i_t, \ldots, i_{t+q-1}\right). \tag{A.2}$$

If we assume that the process is of Markov type, i.e. the true state at a given time instant only depends on the true state at the previous time instant, this reduces to

$$\begin{aligned} p(\mathbf{i}) &= p(i_t)p(i_{t+1}|i_t)p(i_{t+2}|i_{t+1}) \cdots p\left(i_{t+q}|i_{t+q-1}\right) \\ &= p(i_t)\prod_{j=1}^{q} p\left(i_{t+j}|i_{t+j-1}\right). \end{aligned} \tag{A.3}$$

Assuming a stationary (i.e., time-homogeneous) Markov model where these conditional probabilities do not depend on j, (A.3) can be summarized using the set of initial probabilities $p(i_t)$ and the set of so-called transition probabilities $p(i_t|i_{t-1})$.

Working now similarly on $p(\widehat{\mathbf{i}}|\mathbf{i})$, we can write that

$$\begin{aligned} p\left(\widehat{\mathbf{i}}|\mathbf{i}\right) = p\left(\widehat{i}_t|\mathbf{i}\right) p\left(\widehat{i}_{t+1}|\widehat{i}_t, \mathbf{i}\right) p\left(\widehat{i}_{t+2}|\widehat{i}_t, \widehat{i}_{t+1}, \mathbf{i}\right) \\ \ldots p\left(\widehat{i}_{t+q}|\widehat{i}_t, \ldots, \widehat{i}_{t+q-1}, \mathbf{i}\right). \end{aligned} \tag{A.4}$$

Clearly, knowing the attributed class at the previous time instants does not provide any additional information if we already know the true class at all time instants, so (A.4) simplifies to

$$p\left(\widehat{\mathbf{i}}|\mathbf{i}\right) = p\left(\widehat{i}_t|\mathbf{i}\right) p\left(\widehat{i}_{t+1}|\mathbf{i}\right) p\left(\widehat{i}_{t+2}|\mathbf{i}\right) \ldots p\left(\widehat{i}_{t+q}|\mathbf{i}\right). \tag{A.5}$$

Moreover, knowing the true class at all time instant does not provide any additional information if we already know the true class at the same time instant, so (A.5) simplifies further to

$$\begin{aligned} p\left(\widehat{\mathbf{i}}|\mathbf{i}\right) &= p\left(\widehat{i}_t|i_t\right) p\left(\widehat{i}_{t+1}|i_{t+1}\right) p\left(\widehat{i}_{t+2}|i_{t+2}\right) \ldots p\left(\widehat{i}_{t+q}|i_{t+q}\right) \\ &= \prod_{j=0}^{q} p\left(\widehat{i}_{t+j}|i_{t+j}\right). \end{aligned} \tag{A.6}$$

If these conditional probabilities do not change with j, (A.6) can be summarized using the set of so-called emission probabilities $p(\widehat{i}_t|i_t)$. Plugging now (A.3) and (A.6) into (A.1) gives the result stated in (1).

ACKNOWLEDGMENT

The authors would like to thank T. De Maet and J. Radoux for processing the image land cover classifications and Julien's advice on validation.

REFERENCES

[1] K. K. Goldewijk, A. Beusen, G. Van Drecht, and M. De Vos, "The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years," *Global Ecol. Biogeogr.*, vol. 20, no. 1, pp. 73–86, 2011.

[2] W. Li *et al.*, "Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015)," *Earth Syst. Sci. Data*, vol. 10, no. 1, pp. 219–234, 2018.

[3] J.-P. Malingreau and C. J. Tucker, "Large-scale deforestation in the southeastern Amazon basin of Brazil," *Ambio*, vol. 17, no. 1, pp. 49–55, 1988.

[4] G. Duveiller, P. Defourny, B. Desclée, and P. Mayaux, "Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 1969–1981, May 2008.

[5] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, 2013.

[6] K. C. Seto, M. Fragkias, B. Güneralp, and M. K. Reilly, "A meta-analysis of global urban land expansion," *PLoS ONE*, vol. 6, no. 8, Aug. 2011, Art. no. e23777.

[7] S. Angel, J. Parent, D. L. Civco, A. Blei, and D. Potere, "The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050," *Prog. Planning*, vol. 75, no. 2, pp. 53–107, Feb. 2011.

[8] R. J. Keenan, G. A. Reams, F. Achard, J. V. de Freitas, A. Grainger, and E. Lindquist, "Dynamics of global forest area: Results from the FAO global forest resources assessment 2015," *Forest Ecol. Manage.*, vol. 352, pp. 9–20, Sep. 2015.

[9] M. Ballin, G. Barcaroli, M. Masselli, and M. Scarnò, "Redesign sample for land use/cover area frame survey (LUCAS) 2018," *Eurostat, Stat. Work. Papers*, vol. 10, Oct. 2018, Art. no. 132365.

[10] M. A. Friedl *et al.*, "MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 168–182, Jan. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425709002673

[11] S. Bontemps *et al.*, "Revisiting land cover observation to address the needs of the climate modeling community," *Biogeosciences*, vol. 9, no. 6, pp. 2145–2157, Jun. 2012. [Online]. Available: http://www.biogeosciences.net/9/2145/2012/

[12] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, USA: CRC Press, 2019.

[13] S. P. Abercrombie and M. A. Friedl, "Improving the consistency of multitemporal land cover maps using a hidden Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 703–713, Feb. 2016.

[14] B. Desclée, P. Bogaert, and P. Defourny, "Forest change detection by statistical object-based method," *Remote Sens. Environ.*, vol. 102, nos. 1–2, pp. 1–11, 2006.

[15] H. S. J. Zald *et al.*, "Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and above-ground biomass in Saskatchewan, Canada," *Remote Sens. Environ.*, vol. 176, pp. 188–201, Apr. 2016.

[16] P. Defourny *et al.*, "Land cover climate change initiative-product user guide v2. Issue 2.0," ESA-UCLouvain-Geomatics, Louvain-la-Neuve, Belgium, Tech. Rep., 2017. [Online]. Available: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

[17] Z. Zhu and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," *Remote Sens. Environ.*, vol. 144, pp. 152–171, Mar. 2014.

[18] P. Olofsson, G. M. Foody, S. V. Stehman, and C. E. Woodcock, "Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation," *Remote Sens. Environ.*, vol. 129, pp. 122–131, Feb. 2013.

[19] B. Mor, S. Garhwal, and A. Kumar, "A systematic review of hidden Markov models and their applications," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1429–1448, 202, doi: 10.1007/s11831-020-09422-4.

[20] J. Radoux, C. Lamarche, E. Van Bogaert, S. Bontemps, C. Brockmann, and P. Defourny, "Automated training sample extraction for global land cover mapping," *Remote Sens.*, vol. 6, no. 5, pp. 3965–3987, May 2014.

[21] W. Kuang, J. Liu, J. Dong, W. Chi, and C. Zhang, "The rapid and massive urban and industrial land expansions in China between 1990 and 2010: A CLUD-based analysis of their trajectories, patterns, and drivers," *Landscape Urban Planning*, vol. 145, pp. 21–33, Jan. 2016, doi: 10.1016/j.landurbplan.2015.10.001.

[22] A. H. Strahler *et al.*, "Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps," *Eur. Communities, Luxembourg*, vol. 51, no. 4, pp. 1–60, 2006.

[23] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sens. Environ.*, vol. 148, pp. 42–57, May 2014.

[24] S. Khorram, *Accuracy Assessment of Remote Sensing-Derived Change Detection*. Bethesda, MD, USA: ASPRS Publications, 1999.

[25] W. G. Cochran and G. William, *Sampling Techniques*. New York, NY, USA: Wiley, 1977.

[26] D. Saah *et al.*, "Collect Earth: An online tool for systematic reference data collection in land cover and use applications," *Environ. Model. Softw.*, vol. 118, pp. 166–171, Aug. 2019, doi: 10.1016/j.envsoft.2019.05.004.

[27] S. V. Stehman and G. M. Foody, "Key issues in rigorous accuracy assessment of land cover products," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111199, doi: 10.1016/j.rse.2019.05.018.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.

[28] P. Griffiths, B. Jakimow, and P. Hostert, "Reconstructing long term annual deforestation dynamics in Pará and Mato Grosso using the Landsat archive," *Remote Sens. Environ.*, vol. 216, pp. 497–513, Oct. 2018.
[29] W.-K. Ching, M. K. Ng, and E. S. Fung, "Higher-order multivariate Markov chains and their applications," *Linear Algebra Appl.*, vol. 428, nos. 2–3, pp. 492–507, Jan. 2008.
[30] R. Begleiter, R. El-Yaniv, and G. Yona, "On prediction using variable order Markov models," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 385–421, 2004.
[31] E. Otranto, "A time varying hidden Markov model with latent information," *Stat. Model.*, vol. 8, no. 4, pp. 347–366, Dec. 2008.
[32] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial–contextual information in very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 631–651, Sep. 2013.
[33] M. Pastorino *et al.*, "Multisensor and multiresolution remote sensing image classification through a causal hierarchical Markov framework and decision tree ensembles," *Remote Sens.*, vol. 13, no. 5, p. 849, Feb. 2021.



Patrick Bogaert received the Ph.D. degree in agricultural engineering from the Universite Catholique de Louvain (UCLouvain), Leuven-la-Neuve, Belgium, in 1997.

He is currently a Full Professor with UCLouvain and also leads the Environmetrics Research Laboratory with a special focus on spatial statistics and environmental modeling in the broad sense. His research activities also aim to develop efficient methods for processing space-time data in the framework of multisource information, with applications in water resources, soil sciences, and remote sensing at various spatial and temporal scales.



Céline Lamarche received the M.S. degree in bioscience engineering from the Université catholique de Louvain (UCLouvain), Leuven-la-Neuve, Belgium, in 2009.

She is currently a Research Scientist with the Earth and Life Institute, UCLouvain, where she has contributed to several international projects related to land cover mapping at the global and regional scales, such as the medium and high land cover components of the ESA Climate Change Initiative and EU Copernicus Climate Change projects for land cover. Her research interests include optical remote sensing, vegetation monitoring, land use, and land-cover change characterization in the context of climate change.



Pierre Defourny (Member, IEEE) received the Ph.D. degree in agricultural engineering from the Universite Catholique de Louvain (UCLouvain), Leuven-la-Neuve, Belgium, in 1992.

He spent several years in remote sensing research in Africa, Southeast Asia, and Central America. He was a Founding President with the Earth and Life Institute, UCLouvain, where he is currently a Full Professor. He leads the Geomatics Research Laboratory with a special focus on open source system development to exploit Sentinel missions for agriculture and forestry monitoring. His research activities also aim at developing monitoring systems of land cover/land-use change across spatial and temporal scales.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on February 16,2022 at 11:29:12 UTC from IEEE Xplore. Restrictions apply.