

REVEALING LONG-TERM PHYSIOLOGICAL TRAJECTORIES OF GRASSLANDS FROM LEGACY B&W AERIAL PHOTOGRAPHS

Q. Poterek^{1,*}, P.-A. Herrault¹, G. Forestier², D. Schwartz¹

¹ LIVE, CNRS UMR 7362, University of Strasbourg, 67083 Strasbourg, France -
(quentin.poterek, pierre-alexis.herrault, dominique.schwartz)@live-cnrs.unistra.fr

² IRIMAS, University of Haute-Alsace, 68093 Mulhouse, France -
germain.forestier@uha.fr

Commission III, WG 10

KEY WORDS: Archive Aerial Photographs, Radiometric Correction, Colorization, Digital Greenness Metrics, Phenology, Time Series

ABSTRACT:

Landscape reconstruction is crucial to measure the effects of climate change or past land use on current biodiversity. In particular, retracing past phenological changes can serve as a basis for explaining current patterns of plant communities and predict the future extinction of species. Old spatial data are currently used to reconstruct vegetation changes, both morphologically (with landscape metrics) and semantically (grasslands to crops for instance). However, poor radiometric properties (single panchromatic channel, illumination variation, etc.) do not offer the possibility to compute environmental variables (e.g. NDVI and color indices), which strongly limits long-term phenological reconstruction. In this study, we propose a workflow for reconstructing phenological trajectories of grasslands from 1958 to 2011, in the French central Vosges, from old aerial black and white (B&W) photographs. Noise and vignetting corruptions were first corrected in B&W photographs with non-local filtering algorithms. Panchromatic scans were then colorized with a Generative Adversarial Network (GAN). Based on the predicted channels, we finally computed digital greenness metrics (Green Chromatic Coordinate, Excess Greenness) to measure vegetation activity in grasslands. Our results demonstrated the feasibility of reconstructing long-term phenological trajectories from legacy photographs with insights at different levels : (1) the proposed correction methods provided radiometric improvements in old aerial missions; (2) the colorization process led to promising and plausible colorized historical products; (3) digital greenness metrics were useful for describing past vegetation activity.

1. INTRODUCTION

It is widely recognized that present day biodiversity may reflect past land use or past climate (Jansson, Davies, 2007) because of a possible delay in the response of certain species to habitat perturbations (Kuussaari et al., 2009). This interval depends on changes themselves (e.g. intensity and cyclicity) as well as species traits, namely mobility dispersion abilities. In order to assess time-lag between changes in a landscape, and changes in the associated populations, studies rely most of the time on old spatial data that help highlighting spatio-temporal trajectories over large extents and long time periods (Proença et al., 2017).

Among these spatial data, legacy aerial photographs are largely under-exploited while they offer unique opportunities to monitor landscapes at a very high spatial resolution ($\leq 1\text{m}$) from up to the early 1930s (Morgan et al., 2010, Morgan et al., 2017). Despite these advantages, their use remains problematic for many reasons. First, they feature heterogeneous specifications and quality (e.g. noise and vignetting), mainly due to the properties of the acquisition system and the scanning procedure. Second, they lack inherent exploitable attributes, especially in the case of panchromatic pictures, making the development of automatic processing chains a complex task (Paine, Kiser, 2012, Aber et al., 2016).

Another major problem refers to the type of information that can be retrieved from these data. Landscape reconstruction

is generally carried out one of two different ways : (1) morphologically by monitoring landscape metrics (e.g. area and connectivity) over the considered time period (Franco, Morgan, 2007, Herrault et al., 2015); (2) semantically by describing changes in the nature of spatial objects (grasslands to crops for instance) (Treitz, Rogan, 2004). Recently, analysis of radiometric properties in aerial photographs provided useful information to monitor vegetation health or forest phenology (Franco, Morgan, 2007, Reid et al., 2016). Indeed, metrics such as the Green Chromatic Coordinate (GCC) or Excess Greenness (EG) are invariant to illumination conditions and outperform conventional indices such as the NDVI when processing near-sensing images (Nijland et al., 2014). Unfortunately, inherent characteristics of the oldest aerial missions (i.e. B&W photographs) do not offer the possibility to retrieve these indices (Morgan et al., 2010). Thus, long term phenological reconstructions are strongly restricted temporally.

In this study, we develop a generic framework to reconstruct phenological trajectories of grasslands from 1958 to 2011, in the central Vosges from old aerial B&W photographs. We propose a workflow divided in three major steps :

1. Correction of noise and vignetting effects in old aerial B&W photographs;
2. Spectralization of old aerial B&W photographs;
3. Reconstruction of changes in grasslands with greenness indices.

* Corresponding author

2. STUDY SITE

The study site is located in the central Vosges, 30 km from Strasbourg in the East of France (Figure 1). It covers approximately 15,000 ha and includes 10 municipalities. This is a hilly region (100 to 400m) dissected by north-south valleys. The climate is semi-continental with oceanic influences. It comprises a majority of grasslands and crops (wheat, corn) with remnant traditional orchards delineated by hedges. A positive gradual landscape openness is directly observable from North to South, indicating changes in management practices, from traditional to intensive agriculture. Thus, heterogeneous landscape trajectories were provoked by this management diversity and might explain part of the spatial distribution of grasslands plant species over the study site.

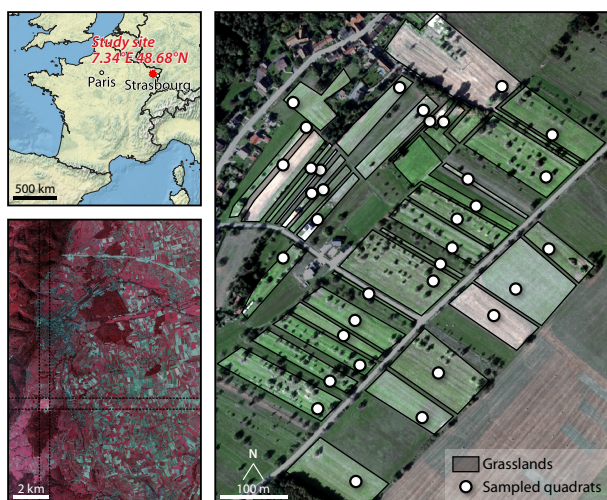


Figure 1. Study site location and surroundings. Infrared false color (bottom left panel) and natural color (right panel) images are provided for visualization.

In the scope of this paper, 34 grasslands were selected to be monitored according to two criteria: (1) grasslands must persist without discontinuity over time, and (2) they vary in terms of openness throughout the years.

3. METHODOLOGY

3.1 Preparation of spatial data

Six air missions acquired between 1958 and 2011 were used to reconstruct long term phenological trajectories of grasslands (Table 1). For testing out the proposed methodology, only 1 single photograph was included in our database for each mission. All photographs that were selected overlapped. Each picture was acquired in spring (from May to June). This means that the proposed photographs captured the signal of grasslands just before the first peak of annual greenness or after the first mowing event of the year. Therefore, for specific years, some grasslands exhibited an almost bare soil.

Panchromatic scans were manually georeferenced and rectified using around 10 GCPs and by applying a second-degree polynomial transformation between image and ground coordinates ($RMSE = 1.21 \pm 0.33$ m). The most recent RGB orthophoto-graph acquired in 2011 systematically served as a reference for the registration of B&W aerial photographs.

Date	Scale/Pixel size(m)	Bands	Type	Provider
01/05/1958	1:24,613 / 0.94*	Pan	Scanned film	IGN
29/05/1976	1:19,707 / 0.40*	Pan	Scanned film	IGN
29/06/1986	1:16,918 / 0.34*	Pan	Scanned film	IGN
29/06/1992	1:28,108 / 0.62*	Pan	Scanned film	IGN
08/06/2000	1:30,995 / 0.60*	Pan	Scanned film	IGN
05/05/2011	- / 0.16	RGB	Digital	CIGAL

Table 1. Specifications for the selected aerial photographs. * For scanned films, scale is followed by a pixel size estimate that was computed using calibration certificates.

In order to demonstrate the limitations imposed by historical photos, as well as the techniques available to address them, a processing chain has been developed to reconstruct long-term phenological trajectories (Figure 2). The workflow can be divided in three steps : (1) correction and harmonization of old aerial photographs, (2) spectralization of panchromatic photographs and (3) phenological time series analysis.

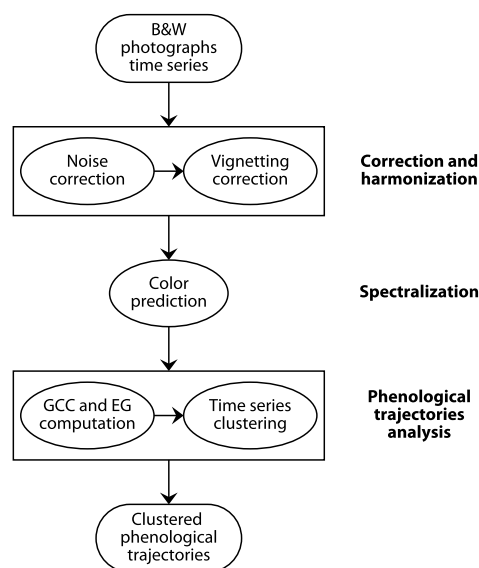


Figure 2. Proposed workflow for reconstructing phenological trajectories of grasslands from old aerial photographs

3.2 Correction and Harmonization of spatial data

3.2.1 Noise correction Noise in aerial photographs is the result of data acquisition or transmission (Jalobeanu et al., 2002). The consequence is a random variation in pixels values, independent of the original data, making interpretation and processing complicated tasks (Jalobeanu et al., 2002, Corner et al., 2003). In the context of old aerial photographs, Gaussian corruption is the most prominent, mainly due to illumination, sensor temperature and film scanning. Since aerial stills were not acquired following a standard and non-evolving scheme, the degree of corruption may vary across the entire time series (Morgan et al., 2010). Thus, noise analysis and correction are mandatory for harmonizing the available photographs.

In this paper, we assess noise with a pyramid-based method. The standard deviation of the Gaussian noise distribution was first obtained by computing the median absolute deviation of the wavelet detail coefficients (Donoho, Johnstone, 1994) at 6 different scales. Each photograph in the series was resampled into lower resolution overviews, by factors of 1, 2, 3, 4, 6 and 12. The outputs were then averaged for each scale and year so as to retrieve an estimate of the standard deviation for each photograph.

After estimation, noise in photographs was corrected with the non-local means (NLM) algorithm (Buades et al., 2011). This technique has been used in a significant number of remote sensing applications, where it was able to smooth surfaces while retaining fine details, such as roads and edges, as in (Huang et al., 2017) for example. Unlike other filtering techniques, NLM performs denoising at patch-level. The algorithm recursively takes one pixel to process, along with its neighborhood. It then retrieves all the patches available in the image and measures their similarity to the processed area. An average of the patch values is then calculated, each contribution being weighted by the degree of similarity to the processed pixel. The standard deviation estimates that we previously computed were passed to the algorithm so as to guide the weighting process. Finally, the pixel value is updated with the weighted and averaged contributions.

3.2.2 Vignetting correction In aerial photographs, spatially organized variations in brightness can be observed, most of the time following a radial gradient (Yu et al., 2004). This results in a bright principal point and dark borders around the fiducial marks. This phenomenon is called vignetting and can be troublesome when the photographs in question are used in photogrammetry or remote sensing applications. Indeed, the gradual darkening of the image results in a non-homogeneous signal for similar surfaces. This can notably disturb the matching of homologous points for the relative orientation of several photographs (Kim, Pollefeys, 2008), reduce image classification performance, or compromise the analysis of time series (Kelcey, Lucieer, 2012).

We propose a simple but efficient technique for removing not only vignetting, but also other local variations in brightness. After masking fiducial marks, the estimation of a correction map was carried out following a pyramid scheme similar to the one previously described for measuring the standard deviation of Gaussian noise. Each photograph was resampled into lower resolutions. All scales were sized back to the original photograph dimensions using bicubic interpolation, and later averaged, in order to compute a multi-scale mean illumination map. It was then subtracted from the initial still, thus eliminating any illumination effect, whether due to exposure or vignetting for example. Finally, pixel values from the output were rescaled so as to match the distribution of the original.

3.2.3 Spectralization of B&W photographs After correction and in order to ensure spatial coherence, all of the available photos were resampled to a 1m resolution. It was then necessary to enrich the spectral content of the available series before calculating phenological metrics and characterizing vegetation activity at each date. Indeed, panchromatic and single-infrared channels provided by old aerial missions are insufficient for the calculation of environmental variables (NDVI or color indices). Thus, we recommend the development of a spectralization technique that would help predict new channels in the visible domain, based on initial panchromatic information.

In essence, spectralization corresponds to colorization. Several techniques are available for colorizing black and white photographs. Recently, various techniques based on deep learning have been proposed, mostly non-convolutional networks (Cheng et al., 2015), CNNs (Zhang et al., 2016) and GANs (Isola et al., 2017). Unlike other methods that rely on human intervention for placing color scribbles (Yatziv, Sapiro, 2006), guiding transfer algorithms (Welsh et al., 2002) or computing appropriate attributes for standard machine learning (Desh-

pande et al., 2015), the advent of deep learning has helped alleviate the colorization process, as models automatically learn feature maps to solve the grayscale-to-color mapping problem.

In the scope of this work, we propose a deep colorization model based on a generative adversarial network (GAN). It is built around two neural networks pitted against one another. The first corresponds to a generator G that learns how to produce samples that could belong to a reference distribution. The second corresponds to a discriminator D , whose role is to distinguish true samples from generated ones (Goodfellow et al., 2014). The end objective is to train a generator capable of deceiving the discriminator. In this paper, we conditioned G by giving it grayscale samples as inputs, so as to learn color channels (Mirza, Osindero, 2014). It should be pointed out that, as of the submission of this paper, the colorization of aerial photographs time series has not been tackled in any other work. The only other papers that come close were either focused on processing single-date products, with (1) the colorization of single-polarization SAR images to obtain full-polarization samples (Song et al., 2018), or (2) the colorization of modern high resolution multispectral satellite imagery (Liu et al., 2018). Thus, the proposed methodology is meant to address different shortcomings, with the colorization of archive spatial data that were computed into long-term time series.

Since there is no proper aerial photographs database available at the moment for deep learning development, we built our own training and validation data sets. To minimize training time, and since the aim was only to colorize old photographs for a small extent, we randomly extracted a total of 3,000 128×128 samples from the 05/05/2011 color reference, spatially independent from one another. Samples were split into training and validation sets given a 10:1 ratio.

In order to learn a panchromatic-to-color mapping, all samples were converted from the RGB color space to CIE Lab. This resulted in images with 3 channels: L , a and b . L contains luminance information, and approximately corresponds to a black and white photograph. Its values range from 0 to 100. The last two channels, a and b , are uncorrelated and contain color information. Their values range from -128 to $+127$. This step allowed separating grayscale and chromatic information. Thus, the proposed model only had to learn how to predict a and b , while preserving spatial information provided by L after concatenation of the input and the two predicted channels. This step also made it possible to learn only two channels (a and b) instead of three (R , G and B), making the training step less costly in terms of time and resources. At last, pixel values from the Lab samples were scaled to a $[-1; +1]$ range as a prerequisite for traditional GAN training (Goodfellow et al., 2014).

The generator was based on a fully-convolutional UNet architecture (Ronneberger et al., 2015). It was fed with a batch of black and white (L) samples that were first downsampled, and later upsampled after passing through a bottleneck. The output corresponded to a and b channels for the input samples. Features learned during downsampling were concatenated in a symmetric fashion to their corresponding upsampled counterparts. This technique helped recognize spatial semantics at various scales, while maintaining spatial information. During training, samples passed to the generator were augmented by applying random horizontal and vertical flips, rotation, blur and variation in lightness. Data augmentation was necessary for simulating the look and feel of old aerial photographs taken under different conditions with specific sensors. In the end, the

generator was trained to minimize the probability of the discriminator labeling generated data as not being part of the reference color distribution (Goodfellow et al., 2014). The cost function for the generator was defined as:

$$L_G = E [\log(D(G(p), p))] \quad (1)$$

where G = generator
 D = discriminator
 p = panchromatic sample
 $G(p)$ = colorized sample

The discriminator was based on the PatchGAN architecture proposed by (Isola et al., 2017). It was fed alternatively with real color and colorized samples and was tasked with the evaluation of generated chroma, in regards to the reference distribution. The discriminator was trained to maximize its probability of distinguishing between both real and generated samples (Goodfellow et al., 2014). The cost function for the discriminator was defined as:

$$L_D = E [\log(D(c, p))] + E [\log(1 - D(G(p), p))] \quad (2)$$

where D = discriminator
 G = generator
 p = panchromatic sample
 c = color sample
 $G(p)$ = colorized sample

The model was trained for 1,000 epochs, using a free Tesla K80 GPU provided by the Google Colab platform.

Colorized photographs from the validation set were evaluated with image quality metrics such as peak signal to-noise ratio (PSNR), and structural similarity index (SSIM). PSNR provides information regarding the quality of reconstruction after colorization. Its values range from 0 db (different) to $+\infty$ (identical). The SSIM was computed in order to provide information regarding the similarity in lightness, contrast and color fidelity. Unlike PSNR, it is said to correlate with human vision assessment (Wang et al., 2004). Its values range from 0 (different) to 1 (identical).

Finally, after training and evaluation, the old aerial photographs (L) were colorized using the generator. The input L channel, along with the predicted a and b channels, were concatenated and converted to RGB, in order to compute phenological indices.

3.3 Time series analysis

Based on the corrected and harmonized photographs, we were then able to analyze changes for grasslands since 1958. It is important to note that the reconstructed trajectories cannot be used to evaluate long-term productivity trends because of irregular acquisition dates between air missions, and the use of one single date per year. Our objective was to assess the possibility of obtaining plausible trajectories after correction and spectralization of old aerial photographs.

First, one 10×10 m quadrat was randomly placed in each of the 34 selected grasslands. All pixels entirely contained in

each quadrat were extracted in order to compute phenological indices.

Then, two digital phenological metrics were calculated. Recent works showed the usefulness of computing color indices for measuring vegetation activity in near-sensing images (Reid et al., 2016). Differences in scene illumination is a major issue when processing old aerial photographs, mainly due to vignetting and variations in exposition (Aber et al., 2016). Taking this issue into account, two metrics were computed in the scope of this study: (1) Green Chromatic Coordinate (GCC, see Equation 3), which is well suited for suppressing variability in scene illumination, and (2) Excess Greenness (EG, see Equation 4), which provides better differentiation between plant material and soil background compared to other color indices (Nijland et al., 2014, Reid et al., 2016).

$$GCC_i = \frac{G_i}{(R_i + G_i + B_i)} \quad (3)$$

$$EG_i = 2 \times G_i - (R_i + B_i) \quad (4)$$

where G_i , R_i and B_i are the predicted green, red and blue pixel values respectively.

In order to reconstruct greenness trajectories from 1958 to 2011, the mean and standard deviation of GCC and EG were computed for each quadrat and year. We also reported these metrics both for the original and the corrected photograph at each date to assess the effects of the proposed correction workflow on the greenness metrics.

Temporal profiles for greenness were finally clustered so as to highlight different groups of grasslands trajectories over time. An agglomerative clustering technique was used. The number of clusters n was set empirically after trying out multiple values. Samples were then merged recursively using cosine distance affinity. More robust time series clustering techniques, such as the ones based on dynamic time warping (Berndt, Clifford, 1994), were not explored in this study due to non-monotonic intervals between air missions (Petitjean et al., 2014).

4. RESULTS AND DISCUSSION

In this section, we first demonstrate the effectiveness of the proposed methodology for correcting and harmonizing old aerial photographs. Results from the pipeline are presented for 1958 in Figure 3.

For all available samples in the time series, we were able to estimate standard deviation of the Gaussian noise corruption. The estimated values based on a 1 byte pixel depth were the following for each year: 1.18 (1958), 3.01 (1976), 1.79 (1986), 2.06 (1992) and 2.44 (2000). Standard deviations were then used to denoise all of the available photographs, based on the NLM algorithm. A visual inspection of the corrected sample shows this technique has managed to retain details, while smoothing the provided photograph. This is further highlighted by the corresponding error map, which is highly textured and contains the removed noisy component (Figure 3). Even though there is no literature regarding the correction of additive Gaussian noise in archive photographs, other papers have come to the same findings. Indeed, the NLM algorithm was successfully applied in

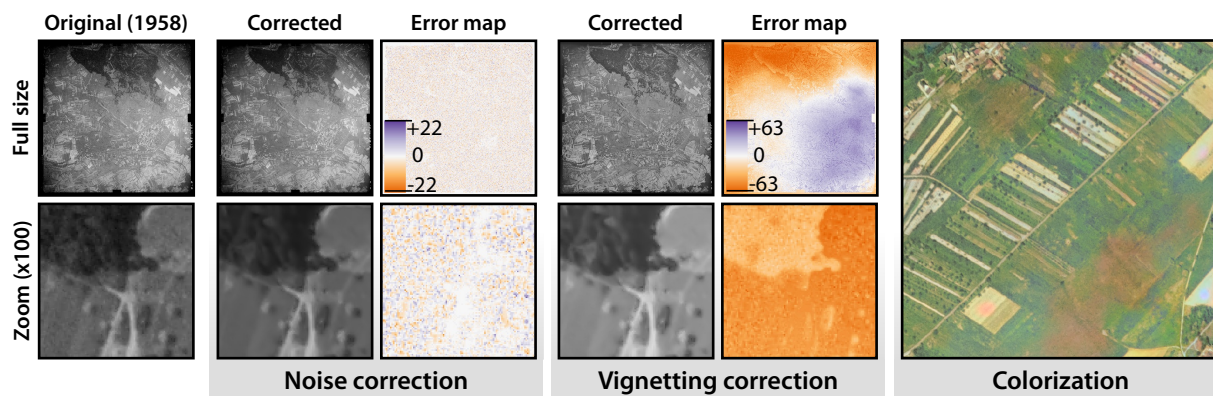


Figure 3. Step-by-step results for the proposed correction and harmonization pipeline. Results are given for the 1958 photograph, as it was the most deteriorated. The original photograph and correction results are presented so as to show the contribution of each step. Error map are also provided, and correspond to the difference between the original and the corrected picture. Finally the colorization result is presented for the studied area.

both spatial (Zhao et al., 2010) and wavelet (Iqbal et al., 2012, Kang et al., 2015) domains, and was able to remove noise while preserving edges and texture, at the cost of some blurring.

The proposed methodology for correcting local and non-local variations in brightness also proved to be efficient. In Figure 3, vignetting is indicated in the original photograph by a strong intensity fall-off around the edges. Meanwhile, the brightest area does not correspond to the center of the picture, nor to its principal point, as would normally be the case. The estimation of a multi-scale mean illumination map made it possible to account for these specific features, as can be seen in the corresponding error map. In an ideal case, the proposed method should only provide an illumination field. However, the error map shows that it also takes into account large spatial structures, such as forests and hedges. One technique to consider for estimating only vignetting, for example, would be to estimate it directly using a time series. The sample to be corrected would be compared to a reference photograph, taken at an earlier or later date. However, the difference between the two pictures should be small enough to minimize possible differences, such as changes in land use, which could affect the estimation of vignetting. Even though the proposed technique was not the same as ours, (da Silva, Candeias, 2012) have obtained similar results, with an increase in mean brightness value and a decrease in standard deviation after vignetting correction.

The distribution of pixel values for each year is shown in Figure 4. Mean pixel values along with standard deviations are given in Table 2 for both the original and corrected samples. Despite visually noticeable changes before and after correction on the photographs (Figure 3), the distribution of pixel values varied only slightly (Figure 4 and Table 2). This is encouraging, as the objective is to preserve the initial signal, while cleaning it of any imperfection that might disturb subsequent analyses. From 1976 to 2000, the mean of pixel values increased slightly, by an average of 0.0525 points. The standard deviation decreased by 0.4425 points on average for the same dates. Similar trends are observed for the 1958 photograph, but with a greater amplitude, as the mean of pixel values increased by 7.67 points, and the standard deviation decreased by 0.9 points. This picture was the most deteriorated, especially due to a strong vignetting, which explains the significance of the corrections that were made.

After correction, photographs were colorized using a generative

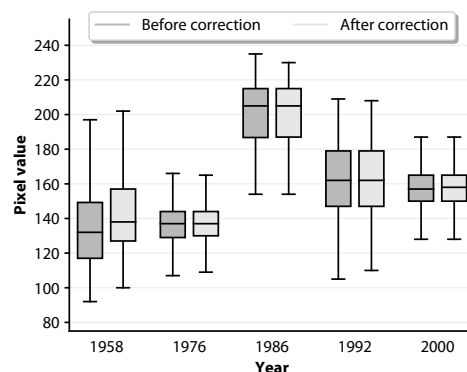


Figure 4. Distribution of pixel values, before and after correction, for each of the studied years.

Date	Before correction	After correction
1958	137.54 ± 26.83	145.21 ± 25.93
1976	137.25 ± 12.42	137.26 ± 11.98
1986	200.72 ± 16.65	200.76 ± 16.49
1992	163.16 ± 18.22	163.25 ± 18.06
2000	157.16 ± 11.39	157.26 ± 10.38

Table 2. Average and standard deviation of pixel values, before and after correction, for each of the studied years.

adversarial network that was conditioned and pre-trained with a 2011 color reference. After 1000 epochs, results from the validation data set indicate robust performances for the colorization task, with $\overline{PSNR} = 39.17 \pm 4.79$ and $\overline{SSIM} = 0.93 \pm 0.05$. Our model performs as well as or better than approaches proposed in other deep learning articles. In comparison, (Varga, Szirányi, 2017) and (Deshpande et al., 2017) obtained mean SSIM values of 0.89 and 0.93 respectively. Regarding PSNR, (Larsson et al., 2016) and (Liu et al., 2018) obtained scores of 24.45 db and 25.05 db respectively. After training, chroma was predicted for all of the available black and white photographs. A colorization example is shown in Figure 3 for the 1958 sample. No color reference is available for old photographs. However, visual analysis of the results testify to the model's ability to process old photographs, although it was trained with recent pictures. Despite local errors, colors were properly predicted for all types of surfaces, including urban

areas, crops, grasslands and bare soils (Figure 5). Most errors are located around image corners, or come in the form of large brown, blue or cyan patches. This behavior could imply that the model is sensitive to edge effect, and does not perform well over homogeneous surfaces larger than training images. Increasing the height and width of training samples could thus help overcome this issue.

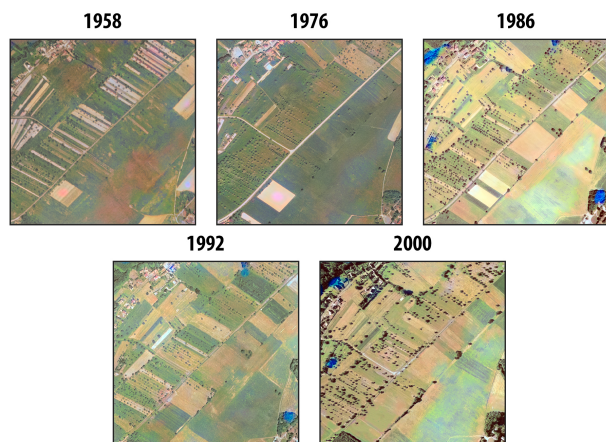


Figure 5. Colorization results for the corrected archive aerial photographs.

Given the decent performances of the model and the adequacy of the colorizations, the latter served as a medium for the calculation of phenological metrics.

We first observe similar relative greenness trends calculated before and after correction (Figure 6). It was an expected result since brightness correction was not meant to drastically change the radiometric distribution of aerial photographs, but rather correct local and global anomalies. Then, mean and standard deviations at each date showed comparable absolute results before and after correction (Table 3). Differences between GCC and EG means are lower than 0.05 between the two steps, indicating minor effects from brightness correction on the calculation of phenological metrics. These results might be explained by several reasons. First, considering observations made from Figure 4, noise correction had no major impact on the gray levels distribution except decreasing the amplitude of the distribution values. This effect is particularly noticeable in 1986, 1992 and 2000. Nonetheless, patterns in the texture of grasslands remained heterogeneous even if they were attenuated. The NLM algorithm successfully preserved the radiometric heterogeneity in grasslands since it was not considered Gaussian noise. Second, the proposed metrics for measuring phenology are supposed to be insensitive to variations in illumination. It is particularly true for GCC for which the results after correction were similar to that of the original photographs. Outcomes are more nuanced for EG since a significant decrease in the standard deviation was observed after correction on the 1958 mission (from 0.072 to 0.066) for which photographs were the most corrupted by vignetting and dark outliers (Figure 4). Corrections allowed enhancing low brightness values, which reduced the EG variability among grasslands. Our results confirm findings from (Nijland et al., 2014), who recommend the use of GCC instead of EG in images where strong brightness variations occur.

Secondly, GCC and EG temporal profiles showed similar trends over the studied time period (Figure 6). Average profiles only indicated a reverse trend between 1986 and 1992, where GCC

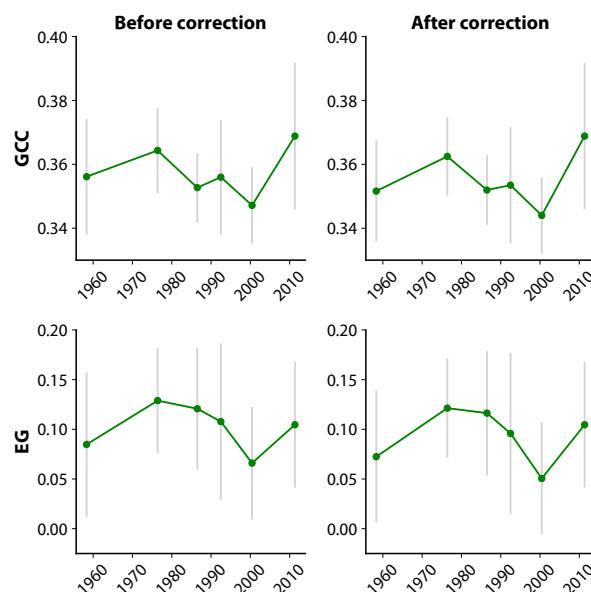


Figure 6. Average temporal profiles of GCC and EG for the sampled grasslands.

Date	Before correction		After correction	
	GCC	EG	GCC	EG
1958	0.36 ± 0.02	0.08 ± 0.08	0.35 ± 0.02	0.07 ± 0.06
1976	0.36 ± 0.01	0.13 ± 0.05	0.36 ± 0.01	0.12 ± 0.05
1986	0.35 ± 0.01	0.12 ± 0.06	0.35 ± 0.01	0.12 ± 0.06
1992	0.36 ± 0.02	0.10 ± 0.08	0.35 ± 0.02	0.09 ± 0.08
2000	0.35 ± 0.01	0.07 ± 0.06	0.34 ± 0.01	0.05 ± 0.06
2011	0.37 ± 0.02	0.11 ± 0.06	—	—

Table 3. Mean and standard deviation for GCC and EG, before and after correction, for each of the studied years.

was increasing (+0.3%) while EG was decreasing (−17%). One explanation might be that EG is more sensitive to extreme values than GCC. Therefore, an excessive number of bare soils in 1992 (19 out of 34 grasslands) contributed to the decrease of mean EG, leading to a trend reversal between those two dates. These results reinforce the necessity of using EG to maximize the differentiation between plant materials and background soils (Nijland et al., 2014, Reid et al., 2016). However, problems might occur in case of non-corrected illuminations variations, typical of the oldest B&W air missions.

High standard deviations were observed for each year, especially for 1958, 1992 and 2011, demonstrating a high greenness variability amongst the grasslands monitored in this study (see Figures 6 and 7). This variability might be explained by several factors. First, species assemblages in grasslands vary strongly, leading to important phenological time-lags between grasslands. Furthermore, recent works showed that the intra-annual phenological cycle is a powerful proxy to predict species diversity in grasslands from satellite image time series (Rapinel et al., 2018, Fauvel et al., 2020). Secondly, the diversity of management practices provoked variations in the timing of mowing events. It results in a large contrast of greenness between grasslands at each date because of the presence of bare soils and grasslands advanced in their phenological cycle.

Last, we tested the possibility of clustering phenological grasslands trajectories (Figure 7). We remind that the retrieved trajectories cannot be used to analyze long-term grasslands productivity because of the non-monotonic acquisition dates, and

the use of one single air mission per year. Nevertheless, clustering results showed the feasibility of identifying distinct groups of grasslands (6 clusters in total) and formulating assumptions. In particular, clusters 1 ($n = 7$) and 3 ($n = 6$) exhibit similar trends over time, but grasslands from the third cluster show higher differences between dates, namely between 1958 and 1976. These grasslands might correspond to ancient hay meadows, with a function that has evolved over time. Individuals from the first cluster showed a higher relative stability, without the presence of bare soils. They might have not been used as a forage resource for a long time. Similarly, clusters 2 ($n = 12$) and 5 ($n = 6$) displayed similar trends, but grasslands from the latter group showed higher greenness values in the most recent photograph. Unlike cluster 2, the species composition of these grasslands might need more time to reach its peak of greenness, leading farmers to mow later in the season.

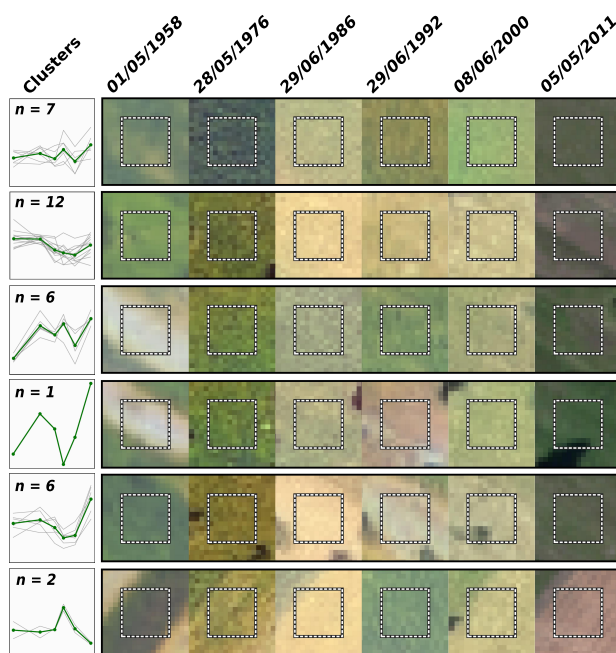


Figure 7. Clusters of trajectories obtained from the GCC on corrected products. In green, the average profile for the cluster. In gray, the individual trajectories for the grasslands. Each square corresponds to one of the 10×10 m quadrats that were used for sampling. Background images correspond to the colorized aerial photographs.

These preliminary results are promising but should be pursued in order to retrace long-term productivity of grasslands. Several options must be considered. First, long-term grasslands trajectories must be retrieved from entire annual cycles to be reconstructed. However, only a handful of air missions are available each year in the best case scenario, with a maximum of 2 to 3 acquisitions for the studied area. One solution would be to densify these series by learning new acquisition dates based on present day satellite time series. Second, these reconstructed phenological series should be analyzed with caution and require further validation due to high uncertainty. *Ex situ* phenological metrics, such as satellite NDVI at lower resolutions, are available from up to early 1980s. *In situ* long-term data, such as herbariums, might also be interesting to analyze species composition and vegetation signal for the oldest available dates.

5. CONCLUSIONS

In this work, we presented a novel methodology for retrieving long-term trends in grasslands phenology. In order to go as far back in time as possible, and to promote the use of national photo libraries, a time series was created with six air missions from 1958 to 2011. Five of which were carried out with panchromatic sensors. Due to the heterogeneity of archive aerial products, the photographs were corrected from noise and vignetting, and then colorized with a conditional GAN to provide consistent multispectral information on the remotely sensed surfaces. The proposed techniques for correction and harmonization showed good performances for the entire series. The colorized photos then served as a basis for the computation of greenness indices, namely GCC and EG, which can be used as a proxy for assessing biodiversity. The output time series showed specific trends for different grasslands, allowing to group them by means of an agglomerative clustering technique. We assume that these clusters reflect different trends in grassland management or species composition in particular. However, the results must be considered with caution, due to the irregular frequency of photo acquisition and the lack of *in situ* and *ex situ* data for validation. Nonetheless, this work demonstrates the usefulness of archive aerial photographs for retrieving long-term and large-scale data for environmental monitoring and evaluating changes in grasslands phenology.

ACKNOWLEDGEMENTS

We thank the French National Institute of Geographic and Forest Information (IGN) for providing us with the archive aerial photographs, along with the corresponding calibration certificates.

REFERENCES

- Aber, J. S., Marzloff, I., Ries, J., 2016. *Small-Format Aerial Photography: Principles, Techniques and Geoscience Applications*. 1st edn, Elsevier Science.
- Berndt, D. J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, AAAIWS'94*, AAAI Press, 359–370.
- Buades, A., Coll, B., Morel, J.-M., 2011. Non-local means denoising. *Image Processing On Line*, 1, 208–212.
- Cheng, Z., Yang, Q., Sheng, B., 2015. Deep colorization. *Proceedings of the IEEE International Conference on Computer Vision*, 415–423.
- Corner, B., Narayanan, R. M., Reichenbach, S. E., 2003. Noise estimation in remote sensing imagery using data masking. *International Journal of Remote Sensing*, 24(4), 689–702.
- da Silva, D. C., Candeias, A. L. B., 2012. Color Restoration of Aerial Photographs. *Special Applications of Photogrammetry*, 1.
- Deshpande, A., Lu, J., Yeh, M.-C., Jin Chong, M., Forsyth, D., 2017. Learning diverse image colorization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6837–6845.

- Deshpande, A., Rock, J., Forsyth, D., 2015. Learning large-scale automatic image colorization. *Proceedings of the IEEE International Conference on Computer Vision*, 567–575.
- Donoho, D. L., Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425–455.
- Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.-L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sensing of Environment*, 237, 111536.
- Franco, J. A., Morgan, J. W., 2007. Using historical records, aerial photography and dendroecological methods to determine vegetation changes in a grassy woodland since European settlement. *Australian Journal of Botany*, 55(1), 1–9.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 2672–2680.
- Herrault, P.-A., Larrieu, L., Cordier, S., Gimmi, U., Lachat, T., Ouin, A., Sarthou, J.-P., Sheeren, D., 2015. Combined effects of area, connectivity, history and structural heterogeneity of woodlands on the species richness of hoverflies (Diptera: Syrphidae). *Landscape Ecology*, 31(4), 877–893.
- Huang, F., Lan, B., Tao, J., Chen, Y., Tan, X., Feng, J., Ma, Y., 2017. A parallel nonlocal means algorithm for remote sensing image denoising on an Intel Xeon Phi platform. *IEEE Access*, 5, 8559–8567.
- Iqbal, M. Z., Ghafoor, A., Siddiqui, A. M., 2012. Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means. *IEEE geoscience and remote sensing letters*, 10(3), 451–455.
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.
- Jalobeanu, A., Blanc-Féraud, L., Zerubia, J., 2002. Estimation of blur and noise parameters in remote sensing. *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4, IEEE, IV–3580.
- Jansson, R., Davies, T. J., 2007. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters*, 11(2), 173–183.
- Kang, W., Yu, S., Seo, D., Jeong, J., Paik, J., 2015. Push-broom-type very high-resolution satellite sensor data correction using combined wavelet-fourier and multiscale non-local means filtering. *Sensors*, 15(9), 22826–22853.
- Kelcey, J., Lucieer, A., 2012. Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. *Remote Sensing*, 4(5), 1462–1493.
- Kim, S. J., Pollefeys, M., 2008. Robust Radiometric Calibration and Vignetting Correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4), 562–576.
- Kuussaari, M., Bommarco, R., Heikkinen, R. K., Helm, A., Krauss, J., Lindborg, R., Öckinger, E., Pärtel, M., Pino, J., Roda, F. et al., 2009. Extinction debt: a challenge for biodiversity conservation. *Trends in ecology & evolution*, 24(10), 564–571.
- Larsson, G., Maire, M., Shakhnarovich, G., 2016. Learning representations for automatic colorization. *European Conference on Computer Vision*, Springer, 577–593.
- Liu, H., Fu, Z., Han, J., Shao, L., Liu, H., 2018. Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks. *Journal of Visual Communication and Image Representation*, 53, 20–30.
- Mirza, M., Osindero, S., 2014. Conditional Generative Adversarial Nets. *CoRR*, abs/1411.1784. <http://arxiv.org/abs/1411.1784>.
- Morgan, J., Gergel, S., Ankerson, C., Tomscha, S., Sutherland, I., 2017. *Historical Aerial Photography for Landscape Analysis*. Springer New York, New York, NY, 21–40.
- Morgan, J. L., Gergel, S. E., Coops, N. C., 2010. Aerial Photography: A Rapidly Evolving Tool for Ecological Management. *BioScience*, 60(1), 47–59.
- Nijland, W., de Jong, R., de Jong, S. M., Wulder, M. A., Bater, C. W., Coops, N. C., 2014. Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras. *Agricultural and Forest Meteorology*, 184, 98–106.
- Paine, D. P., Kiser, J. D., 2012. *Aerial Photography and Image Interpretation*. Wiley.
- Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., Keogh, E., 2014. Dynamic time warping averaging of time series allows faster and more accurate classification. *2014 IEEE international conference on data mining*, IEEE, 470–479.
- Proença, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R. D. et al., 2017. Global biodiversity monitoring: from data sources to essential biodiversity variables. *Biological Conservation*, 213, 256–263.
- Rapinel, S., Rossignol, N., Hubert-Moy, L., Bouzillé, J.-B., Bonis, A., 2018. Mapping grassland plant communities using a fuzzy approach to address floristic and spectral uncertainty. *Applied Vegetation Science*, 21(4), 678–693.
- Reid, A. M., Chapman, W. K., Prescott, C. E., Nijland, W., 2016. Using excess greenness and green chromatic coordinate colour indices from aerial images to assess lodgepole pine vigour, mortality and disease occurrence. *Forest Ecology and Management*, 374, 146–153.
- Song, Q., Xu, F., Jin, Y., 2018. Radar Image Colorization: Converting Single-Polarization to Fully Polarimetric Using Deep Neural Networks. *IEEE Access*, 6, 1647–1661.
- Treitz, P., Rogan, J., 2004. Remote sensing for mapping and monitoring land-cover and land-use change-an introduction. *Progress in planning*, 61(4), 269–279.
- Varga, D., Szirányi, T., 2017. Twin deep convolutional neural network for example-based image colorization. *International Conference on Computer Analysis of Images and Patterns*, 2, Springer, 184–195.
- Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.

Welsh, T., Ashikhmin, M., Mueller, K., 2002. Transferring Color to Greyscale Images. *ACM Trans. Graph.*, 21(3), 277–280.

Yatziv, L., Sapiro, G., 2006. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing*, 15(5), 1120–1129.

Yu, W., Chung, Y., Soh, J., 2004. Vignetting distortion correction method for high quality digital imaging. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, 3, IEEE, 666–669.

Zhang, R., Isola, P., Efros, A. A., 2016. Colorful image colorization. *European conference on computer vision*, Springer, 649–666.

Zhao, M., Zhang, W., Wang, Z., Hou, Q., 2010. Satellite image deconvolution based on nonlocal means. *Applied optics*, 49(32), 6286–6294.