Check for updates

# Hyperspectral Video Analysis by Motion and Intensity Preprocessing and Subspace Autoencoding

Raffaele Vitale[1]\*, Cyril Ruckebusch[1], Ingunn Burud[2] and Harald Martens[3,4]

[1]Univ. Lille, CNRS, LASIRE (UMR 8516), Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, Lille, France, [2]Faculty of Science and Technology, Norwegian University of Life Sciences, Oslo, Norway, [3]Idletechs AS, Trondheim, Norway, [4]Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway

Hyperspectral imaging has recently gained increasing attention from academic and industrial world due to its capability of providing both spatial and physico-chemical information about the investigated objects. While this analytical approach is experiencing a substantial success and diffusion in very disparate scenarios, far less exploited is the possibility of collecting sequences of hyperspectral images over time for monitoring dynamic scenes. This trend is mainly justified by the fact that these so-called hyperspectral *videos* usually result in BIG DATA sets, requiring TBs of computer memory to be both stored and processed. Clearly, standard chemometric techniques do need to be somehow adapted or expanded to be capable of dealing with such massive amounts of information. In addition, hyperspectral video data are often affected by many different sources of variations in sample chemistry (for example, light absorption effects) and sample physics (light scattering effects) as well as by systematic errors (associated, *e.g.*, to fluctuations in the behaviour of the light source and/or of the camera). Therefore, identifying, disentangling and interpreting all these distinct sources of information represents undoubtedly a challenging task. In view of all these aspects, the present work describes a multivariate hybrid modelling framework for the analysis of hyperspectral videos, which involves spatial, spectral and temporal parametrisations of both known and unknown chemical and physical phenomena underlying complex real-world systems. Such a framework encompasses three different computational steps: 1) motions ongoing within the inspected scene are estimated by optical flow analysis and compensated through IDLE modelling; 2) chemical variations are quantified and separated from physical variations by means of Extended Multiplicative Signal Correction (EMSC); 3) the resulting light scattering and light absorption data are subjected to the On-The-Fly Processing and summarised spectrally, spatially and over time. The developed methodology was here tested on a near-infrared hyperspectral video of a piece of wood undergoing drying. It led to a significant reduction of the size of the original measurements recorded and, at the same time, provided valuable information about systematic variations generated by the phenomena behind the monitored process.

Keywords: hyperspectral videos, motion compensation, IDLE modelling, light scattering, light absorption, extended multiplicative signal correction, on-the-fly processing, BIG measurement DATA

# 1 INTRODUCTION

## 1.1 Hyperspectral Videos

In the last decade, hyperspectral imaging has experienced a significant diffusion mainly because of its capability of providing spatial and physico-chemical information about the systems under study - Hugelier et al. (2020). By returning whole spectra for all scanned pixels, in fact, a hyperspectral image permits to map the distribution of the constituents of the investigated samples all over the inspected field of view. For this reason, the applications of this analytical approach have lately dramatically increased in many domains of interest, like medicine, forensics, geoscience, urban and environmental surveillance and fire detection—Fischer and Kakoulli (2006); Chuvieco and Kasischke (2007); Hay et al. (2011); Matikainen and Karila (2011); Elmasry et al. (2012); Lu and Fei (2014); Silva et al. (2017); Khan et al. (2018); Vitale et al. (2020a).

Nonetheless, although hyperspectral imaging devices have become rather common tools in both academic and industrial chemistry laboratories, they are rarely configured so as to collect series of hyperspectral images over time for dynamic scene monitoring. There are two reasons behind this tendency: first of all, finding a reasonable compromise between spatial and spectral resolution and recording rate is not an easy and straightforward task; second, these so-called hyperspectral *videos* often translate into BIG DATA sets that can hardly be coped with by methodologies commonly resorted to for the analysis of individual hyperspectral images—for instance, Principal Component Analysis (PCA), Pearson (1901); Hotelling (1933), Partial Least Squares regression (PLS), Wold et al. (1983); Martens and Næs (1989), Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), Tauler et al. (1995), and Non-Negative Matrix Factorisation (NNMF), Lawton and Sylvestre (1971); Martens (1979). As an example, one can consider that, when storing hundreds of hyperspectral data arrays, the computer memory load is likely to increase up to the order of magnitude of the TBs. Modern workstations cannot readily handle such massive amounts of information and, therefore, standard chemometric techniques do need to be somehow adapted or extended to be possibly utilised in similar scenarios. Furthermore, hyperspectral video data typically account for various phenomena related to sample physics (*e.g.*, light scattering) and sample chemistry (light absorbance) and can be significantly affected by many different types of systematic errors (like those associated to nuisance fluctuations of the light source and/or the camera). Thus, identifying, disentangling, modelling and interpreting all these distinct sources of variations remains undoubtedly a challenging task.
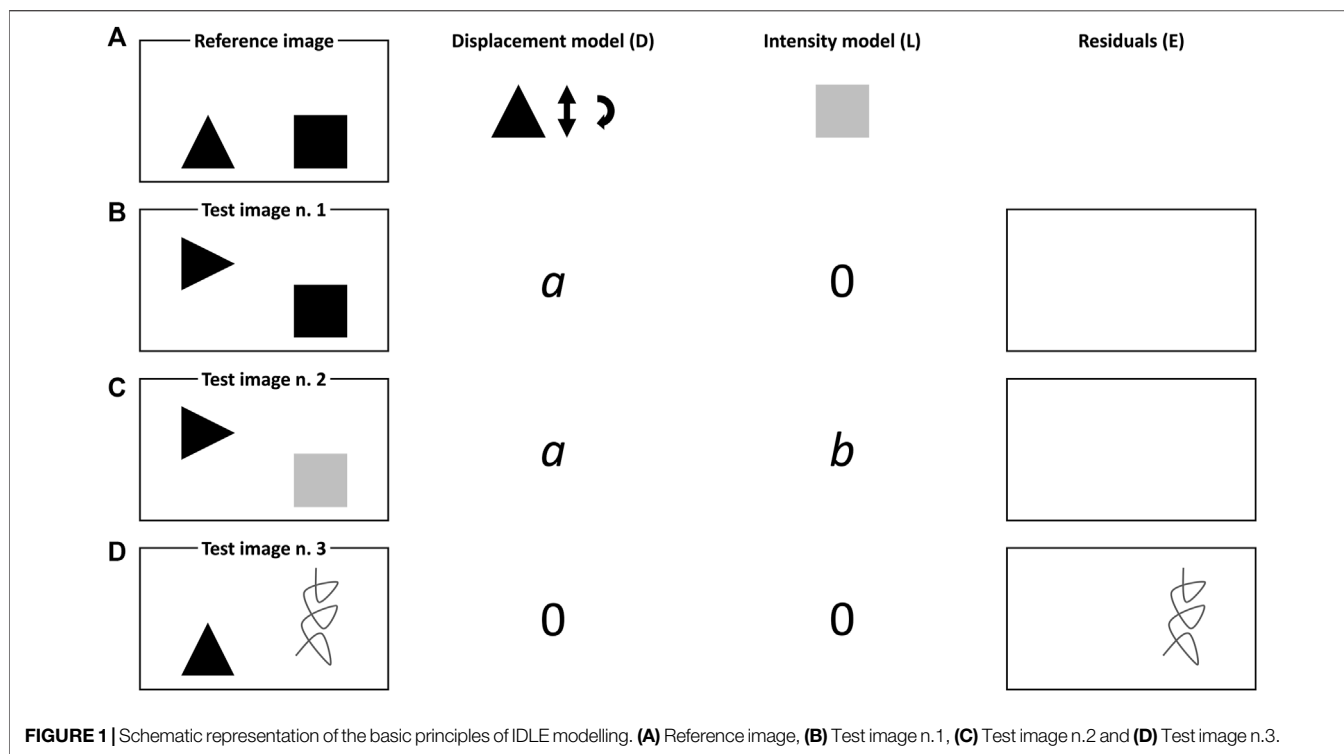
## 1.2 Hyperspectral Video Analysis

Many known causal phenomena influence how light interacts with matter. The most important ones—light absorbance and light scattering—can even be approximated by relatively simple models, *e.g.*, the Beer-Lambert's law or the Kubelka-Munck theory—Bouguer, (1729); Lambert, (1760); Beer (1852); Kubelka and Munk (1931). Yet, albeit some of the

aforementioned error factors affecting spectroscopic measurements—like illumination changes—can also be easily foreseen, a detailed mathematical characterisation of the spectral effects they might generate would be prohibitive from a computational point of view. For this reason, modelling and analysing hyperspectral videos constitutes a problematic challenge. Hyperspectral video data, in fact, yield information about the four main ontological aspects of reality: space, time, properties/attributes (for instance, a light intensity profile in the near-infrared—NIR—spectral range) and their interactions. Thus, a comprehensive description of a hyperspectral video would require the identification and the quantification of *factors* or *components* (both known and unknown) accounting for spatial, temporal and spectral variation patterns in such data. This would allow practitioners and users to gain new insights into complex systems of high relevance and into the interplay between the known and unknown phenomena driving their behaviour and evolution. As an immediate example of such an interplay, consider wood drying, a process exhibiting a deep economic and technical impact—McMillen (1964): water absorption properties allow, in principle, the moisture content of wood samples to be accurately determined. However, these properties may substantially change along with the thermodynamic state of water molecules (*i.e.*, free or bound) and might even mimic spectral contributions from other wood constituents, like cellulose, hemicellulose and lignin. In this as well as in numerous other real-world scenarios, disentangling and characterising the two aforementioned types of phenomena becomes, therefore, crucial from the perspective of understanding. In this article, a novel hybrid approach to achieve this objective is presented. It combines three multivariate approximation strategies for the compression and rational handling of hyperspectral videos: IDLE modelling—Westad and Martens (1999); Martens (2015) — Extended Multiplicative Signal Correction (EMSC)—Martens and Stark (1991); Martens et al. (2003)—and the On-The-Fly Processing (OTFP)—Vitale et al. (2017b). If, on the one hand, EMSC is a well-established tool in the chemometric community, on the other hand IDLE modelling and the OTFP have only recently been conceived, although they have already demonstrated their potential for fast processing of BIG DATA streams—Martens (2015); Vitale et al. (2017b); Stefansson et al. (2020); Vitale et al. (2020b). For the joint analysis of spatial and intensity changes in video recordings, IDLE splits the data variation into two domains as expressed in the following mathematical relation:

$$I = D(L) + E \qquad (1)$$

by which a generic measured image (I) can be described as a function of the displacement (D) of a local intensity image (L) plus error (E). Imagine, for instance, that two different objects (*i.e.*, a black triangle and a black square—see **Figure 1A**) were photographed on a white table. After 1 minute, someone moves the first object along *y*, rotates it 90° and collects another picture (see **Figure 1B**). After 2 minutes, a third picture is taken after the square was painted grey (see **Figure 1C**). Assume now that D and L explain simultaneously vertical displacements and clockwise

**FIGURE 1 |** Schematic representation of the basic principles of IDLE modelling. **(A)** Reference image, **(B)** Test image n.1, **(C)** Test image n.2 and **(D)** Test image n.3.

rotations of the black triangle and variations in the black square pixel intensities, respectively. In this simple illustration, D would encode the triangle movement as an individual coefficient (say, $a$), proportional to the dissimilarity from the object's original location and positioning and exhibiting a sign that depends on the sense of its motion. Analogously, L would quantify the change in the intensity of the pixels of the square as a positive parameter (say, $b$), since the transition from black to light grey implies an increase of the image lightness. Unexpected motions and colours as well as the appearance of unexpected items, like the dark grey spiral-like structure in **Figure 1D**, would be accounted for by the residuals E.

Conversely, the OTFP gradually constructs reduced-rank bilinear models that summarise virtually *ever-lasting* streams of multivariate responses and capture the evolving covariation patterns among their (spectral) variables in space and time. In other words, it represents an extension of classical PCA designed for processing such multivariate responses as soon as they are collected and, most importantly, without requiring entire raw datasets to be kept in memory. More specifically, the OTFP rests on a flexible bilinear subspace model structure which is automatically expanded when a new variation pattern is discovered—as for classical moving-window PCA implementations, Makeig et al. (2000); Wang et al. (2005), even if, here, relevance for *old* or *past* observations is never lost—or refined when the same variation patterns are repeatedly observed, while statistical redundancies are filtered out guaranteeing high rates of information compression. In contrast to black-box *deep learning* solutions, this PCA-like model-based approximation is graphically interpretable in its compressed state and allows at any time the original input to be reconstructed with a better signal-to-noise ratio (as measurement noise is eliminated).

Here, the sequential utilisation of IDLE, EMSC and the OTFP for the investigation of hyperspectral videos will be tested in a context similar to that envisioned before: the monitoring of the drying process of a wood specimen. The results of this study will highlight how this combination can enable an accurate estimation of the dynamic evolution of wood properties and how relatively simple quantitative spatial and temporal information can be extracted from a seemingly overwhelming stream of hyperspectral video data by coupling different mathematical modelling techniques.

## 1.3 Hyperspectral Video Data Structure

Hyperspectral videos can be regarded as time series of three-dimensional data arrays (hyperspectral *frames* or *snapshots*) with dimensions $N_x \times N_y \times J$, where $N_x$ and $N_y$ denote the number of pixels scanned along the horizontal and vertical direction, respectively, and $J$ the number of wavelength channels sampled by the equipment employed. More broadly speaking, they can be thought of as the product of the concatenation of these arrays along a fourth time-related measurement mode. In spite of their multidimensional structure, hyperspectral video data are usually analysed in their unfolded form, *i.e.* as matrices of size $N_x N_y K \times J$ with $K$ representing the amount of time points at which the aforementioned frames are collected. Each row of such matrices carries a single spectral profile recorded for an individual pixel at a given time point.

## 2 METHODS

In this article, EMSC and the OTFP are applied in a sequential fashion to assess/discover and quantify known and unknown

sources of data variability in hyperspectral videos. This strategy combines mechanistic and empirical multivariate modelling for describing all physical, chemical and instrumental variation patterns behind hyperspectral video recordings. In order to account for and compensate possible motions and pixel intensity changes which could originate complex non-linearities distorting the measured spatiospectral response, optical flow analysis—Horn and Schunck (1981, 1993)—and IDLE are applied in a preliminary preprocessing step.

The next sections will describe in detail the basics of the three different methodologies exploited here.

## 2.1 IDLE Modelling

Broadly speaking, the IDLE model is a mathematical description of real-world objects or scenes (characterised by spatiotemporal measurements like videos) in terms of their intensity and spatial variations. Here, IDLE is utilised as an empirical compression approach for sets of consecutive video frames, yielding high compression rates and, at the same time, enabling qualitative and quantitative data interpretation. IDLE is based on a three-step methodological procedure:

1. first of all, it segments out each of the relevant, independent objects (so-called *holons*) within a particular scene;
2. then, for each holon it estimates both D (accounting for motions and shape changes) and L, relative to a fixed, user-defined reference frame;
3. finally, it *morphs back* the holons in the investigated image to their spatial shape and location in the reference image. This facilitates a compact subspace modelling of both displacements and intensity changes.

### 2.1.1 Motion Estimation and Motion Compensation

IDLE modelling concerns how to reduce the complexities that arise when modelling objects that both move and change intensity (or spectral profile) at the same time. Imagine, for instance, a video composed of $K$ grey-scale images ($\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_k, \ldots, \mathbf{I}_K$) of size $N_x \times N_y$ depicting certain objects whose shape and brightness varies over time. Let $\mathbf{I}_{\mathrm{ref}}$ be one of these images, chosen to define a common reference for all the other ones. Analogously, let $\underline{\mathbf{O}}_{\mathrm{ref}}$ ($N_x \times N_y \times 2$) define the horizontal and vertical pixel coordinates (or pixel *addresses*) at which these objects are visible in $\mathbf{I}_{\mathrm{ref}}$. The reference intensity image at pixel adresses $\underline{\mathbf{O}}_{\mathrm{ref}}$ is then $\mathbf{I}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}}$. At this point, the objects in the scene setting captured by each video frame, $\mathbf{I}_k$, can be described with respect to how they look in $\mathbf{I}_{\mathrm{ref}}$. Neglecting motions, at time $k$, the local intensity-corrected version of $\mathbf{I}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}}$ can be expressed as:

$$\mathbf{L}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}} = \mathbf{I}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}} + \Delta\mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}} \tag{2}$$

with $\Delta\mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}$ ($N_x \times N_y$) carrying the image intensity deviations from $\mathbf{I}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}}$.

Likewise, the pixel adresses where the objects from $\mathbf{I}_{\mathrm{ref}}$ are observable in $\mathbf{I}_k$ become:

$$\underline{\mathbf{O}}_k = \underline{\mathbf{O}}_{\mathrm{ref}} + \Delta\underline{\mathbf{O}}_k \tag{3}$$

where $\Delta\underline{\mathbf{O}}_k$ ($N_x \times N_y \times 2$) contain the so-called horizontal and vertical motion fields indicating how every pixel in $\mathbf{I}_k$ should be displaced so that the objects in $\mathbf{I}_k$ mimic their shape and location in $\mathbf{I}_{\mathrm{ref}}$. Hence, merging **Eqs (2)**, **(3)**, the IDLE model for the $k$th frame can be written compactly as a function of how its intensity has changed ($\Delta\mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}$) and how it has moved ($\Delta\underline{\mathbf{O}}_k$) compared to the reference one:

$$\mathbf{I}_{k,\underline{\mathbf{O}}_k} = \mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}+\Delta\underline{\mathbf{O}}_k} = \mathbf{I}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}} + \Delta\mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}} = \mathbf{L}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}} \tag{4}$$

According to this notation, the terms I, D and L in **Eq. (1)** would correspond to $\mathbf{I}_{k,\underline{\mathbf{O}}_k}$, $\Delta\underline{\mathbf{O}}_k$ and $\mathbf{L}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}$, respectively.

From a practical perspective, $\Delta\underline{\mathbf{O}}_k$ can be obtained by motion estimation—Horn and Schunck (1981, 1993)—comparing $\mathbf{I}_k$ and $\mathbf{I}_{\mathrm{ref}}$. This allows one to morph the objects from where they were located in $\mathbf{I}_k$ back to their pixel addresses in $\underline{\mathbf{O}}_{\mathrm{ref}}$ and to their intensity at time $k$ relative to $\mathbf{I}_{\mathrm{ref}}$ ($\mathbf{L}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}$). The intensity changes, $\Delta\mathbf{I}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}$, as well as the motion fields $\Delta\underline{\mathbf{O}}_k$ are all given in the coordinate system of $\mathbf{I}_{\mathrm{ref}}$, *i.e.*, $\underline{\mathbf{O}}_{\mathrm{ref}}$.

### 2.1.2 Dual-Domain Bilinear Modelling of a Hyperspectral Video

Even when a hyperspectral video is handled, all the wavelength channels must follow the same spatial displacement at each time $k$. For this purpose, the unfolded vertical and horizontal motion fields, $\Delta\mathbf{o}_k^{\mathrm{T}}$ ($1 \times 2N_xN_y$), can be estimated from an optimised combination of such channels, gathered column-wise into the matrix $\Delta\mathbf{O}$ ($K \times 2N_xN_y$), modelled bilinearly as

$$\Delta\mathbf{O} = \mathbf{T}_{\Delta\mathbf{O}}\mathbf{P}_{\Delta\mathbf{O}}^{\mathrm{T}} + \mathbf{E}_{\Delta\mathbf{O}}^{\mathrm{T}} \tag{5}$$

and applied to each entire hyperspectral frame. Here, $\mathbf{T}_{\Delta\mathbf{O}}$ ($K \times A_{\mathrm{IDLE}}$) contains the projection coordinates of $\Delta\mathbf{O}$ on the directions defined by the columns of $\mathbf{P}_{\Delta\mathbf{O}}$ ($2N_xN_y \times A_{\mathrm{IDLE}}$) and $\mathbf{E}_{\Delta\mathbf{O}}$ ($K \times 2N_xN_y$) carries the corresponding residuals not explained at the chosen rank, $A_{\mathrm{IDLE}} < 2N_xN_y$.

Compact, low-dimensional bilinear models often summarize quite well the motions in $\Delta\underline{\mathbf{O}}$ when they are defined in the same reference coordinate system. Also the unfolded intensity images $\Delta\mathbf{i}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}^{\mathrm{T}}$ ($1 \times N_xN_y$) may be well approximated in a similar fashion if expressed in a common coordinate system:

$$\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}} = \mathbf{T}_{\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}\mathbf{P}_{\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}} + \mathbf{E}_{\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}} \tag{6}$$

with $\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}$ ($K \times N_xN_y$) being the 2D array resulting from the column-wise concatenation of each $\Delta\mathbf{i}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}}^{\mathrm{T}}$ vector.

Rewriting **Eq. (4)** in vectorial form, the aforementioned morphing operation can be therefore expressed for the $k$th video frame as:

$$\mathbf{i}_{k,\underline{\mathbf{O}}_k}^{\mathrm{T}} = \mathbf{i}_{k,\underline{\mathbf{O}}_{\mathrm{ref}}+\Delta\underline{\mathbf{O}}_k}^{\mathrm{T}} = \mathbf{i}_{k,\mathbf{o}_{\mathrm{ref}}^{\mathrm{T}}+\Delta\mathbf{o}_k^{\mathrm{T}}}^{\mathrm{T}} = \mathbf{i}_{k,\mathbf{o}_{\mathrm{ref}}^{\mathrm{T}}+\left(\mathbf{t}_{k,\Delta\mathbf{O}}^{\mathrm{T}}\mathbf{P}_{\Delta\mathbf{O}}^{\mathrm{T}}+\mathbf{e}_{k,\Delta\mathbf{o}}^{\mathrm{T}}\right)}^{\mathrm{T}}$$

$$= \mathbf{i}_{\mathrm{ref},\underline{\mathbf{O}}_{\mathrm{ref}}}^{\mathrm{T}} + \mathbf{t}_{k,\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}}\mathbf{P}_{\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}} + \mathbf{e}_{k,\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}} \tag{7}$$

where $\mathbf{t}_{k,\Delta\mathbf{O}}^{\mathrm{T}}$, $\mathbf{e}_{k,\Delta\mathbf{O}}^{\mathrm{T}}$, $\mathbf{t}_{k,\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}}$ and $\mathbf{e}_{k,\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}^{\mathrm{T}}$ denote the $k$th row vectors of $\mathbf{T}_{\Delta\mathbf{O}}$, $\mathbf{E}_{\Delta\mathbf{O}}^{\mathrm{T}}$, $\mathbf{T}_{\Delta\mathbf{I}_{\mathrm{ref}}}$ and $\mathbf{E}_{\Delta\mathbf{I}_{\underline{\mathbf{O}}_{\mathrm{ref}}}}$, respectively. Refolding is finally required for the sake of representation.

## 2.2 Extended Multiplicative Signal Correction

EMSC is a bilinear modelling approach that permits to separate, quantify and correct for distinct types of known chemical and physical data variation sources in the acquired signal profiles. As applied in this article, EMSC assumes that a generic spectrum, $\mathbf{x}$ (of dimensions $J \times 1$) can be mathematically described as:

$$\mathbf{x} = b\left(\mathbf{r} + \sum_i \Delta c_i \mathbf{s}_i\right) + a\mathbf{1} + d\mathbf{f} + g\mathbf{f}^2 + \mathbf{e} \qquad (8)$$

where $b$ is the effective relative pathlength; $\mathbf{r}$ ($J \times 1$) is a predetermined reference spectrum; $\Delta c_i$ and $\mathbf{s}_i$ ($J \times 1$) denote the presumed concentration/abundance contribution and the spectral fingerprint of the $i$th main constituent of the system under study, respectively; $\mathbf{1}$ ($J \times 1$) is a column vector of ones; $\mathbf{f}$ ($J \times 1$) contains values monotonically increasing from $-1$ to $1$; $a$, $d$ and $g$ constitute a set of coefficients; and $\mathbf{e}$ ($J \times 1$) carries the unmodelled residuals (*i.e.*, unmodelled chemical and/or physical variations as well as random measurement noise) resulting from this approximation.

Altogether, $\mathbf{1}$, $\mathbf{f}$ and $\mathbf{f}^2$ connote polynomial model dimensions accounting for smoothly wavelength-dependent phenomena (baseline level, slope and curvature, respectively).

Given $h_i = b\Delta c_i$ ($\forall i$), the unknowns in **Eq. (8)** can be retrieved by Ordinary or Weighted Least Squares (OLS/WLS) as:

$$[b \ h_1 \ \ldots \ h_I \ a \ d \ g] = \mathbf{x}^T \mathbf{W}_{\text{EMSC}} \mathbf{W}_{\text{EMSC}} \mathbf{M} \left(\mathbf{M}^T \mathbf{W}_{\text{EMSC}} \mathbf{W}_{\text{EMSC}} \mathbf{M}\right)^{-1} \qquad (9)$$

where $\mathbf{M} = [\mathbf{r} \ \mathbf{s}_1 \ \ldots \ \mathbf{s}_I \ \mathbf{1} \ \mathbf{f} \ \mathbf{f}^2]$ and $\mathbf{W}_{\text{EMSC}}$ ($J \times J$) is a diagonal matrix of weights associated to the different sampled spectral channels[1].

Since the constituent profiles, $\mathbf{s}_i$, are a required input for EMSC processing, this methodology has been chosen for describing expected variation patterns evolving all over the duration of a hyperspectral video.

Once the EMSC coefficients have been calculated as in **Eq. (9)**, they can be exploited for pretreating the input spectrum, $\mathbf{x}$, in order to filter varying light scattering effects as:

$$\mathbf{x}^p = \frac{\left(\mathbf{x} - a\mathbf{1} - d\mathbf{f} - g\mathbf{f}^2\right)}{b} \qquad (10)$$

with $^p$ standing for *preprocessed*. In the present application of EMSC, the estimated chemical variations will also be subtracted from $\mathbf{x}$ as:

$$\mathbf{x}^p = \frac{\left(\mathbf{x} - a\mathbf{1} - d\mathbf{f} - g\mathbf{f}^2 - \sum_i h_i \mathbf{s}_i\right)}{b}$$

$$= \frac{\left(\mathbf{x} - a\mathbf{1} - d\mathbf{f} - g\mathbf{f}^2\right)}{b} - \sum_i \Delta c_i \mathbf{s}_i \qquad (11)$$

Finally, if EMSC residuals are deemed to be affected by the effective optical pathlength of the sample, they can be computed as:

---

[1]If $\text{diag}(\mathbf{W}_{\text{EMSC}}) = \mathbf{1}$, the parameter estimation is carried out by OLS.

$$\mathbf{e} = \mathbf{x} - b\mathbf{r} - \sum_i h_i \mathbf{s}_i - a\mathbf{1} - d\mathbf{f} - g\mathbf{f}^2$$

$$= \mathbf{x} - b\left(\mathbf{r} - \sum_i \Delta c_i \mathbf{s}_i\right) - a\mathbf{1} - d\mathbf{f} - g\mathbf{f}^2 \qquad (12)$$

Pathlength-corrected residuals are subsequently estimated as:

$$\tilde{\mathbf{e}} = b^{-1}\mathbf{e} \qquad (13)$$

## 2.3 The On-The-Fly Processing

After the IDLE-based motion estimation-compensation and the quantification-correction of known physical and chemical variations by EMSC preprocessing, the resulting unmodelled residuals are analysed by the OTFP in the attempt of looking for unknown, yet systematic variability patterns in data. The OTFP relies on a self-learning adaptive modelling principle which allows massive amounts of measurement recordings collected over time to be compressed with a minimal loss of meaningful information according to a PCA-like bilinear decomposition. Its global computational procedure encompasses five different steps:
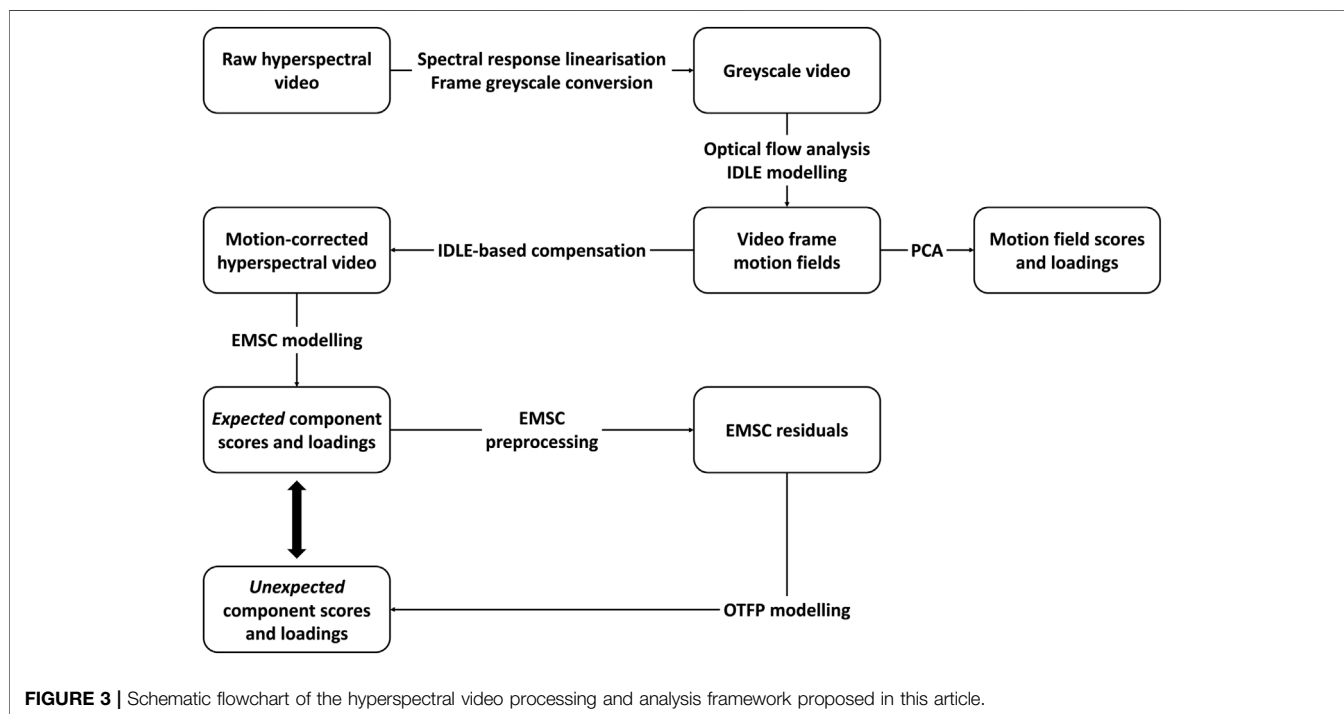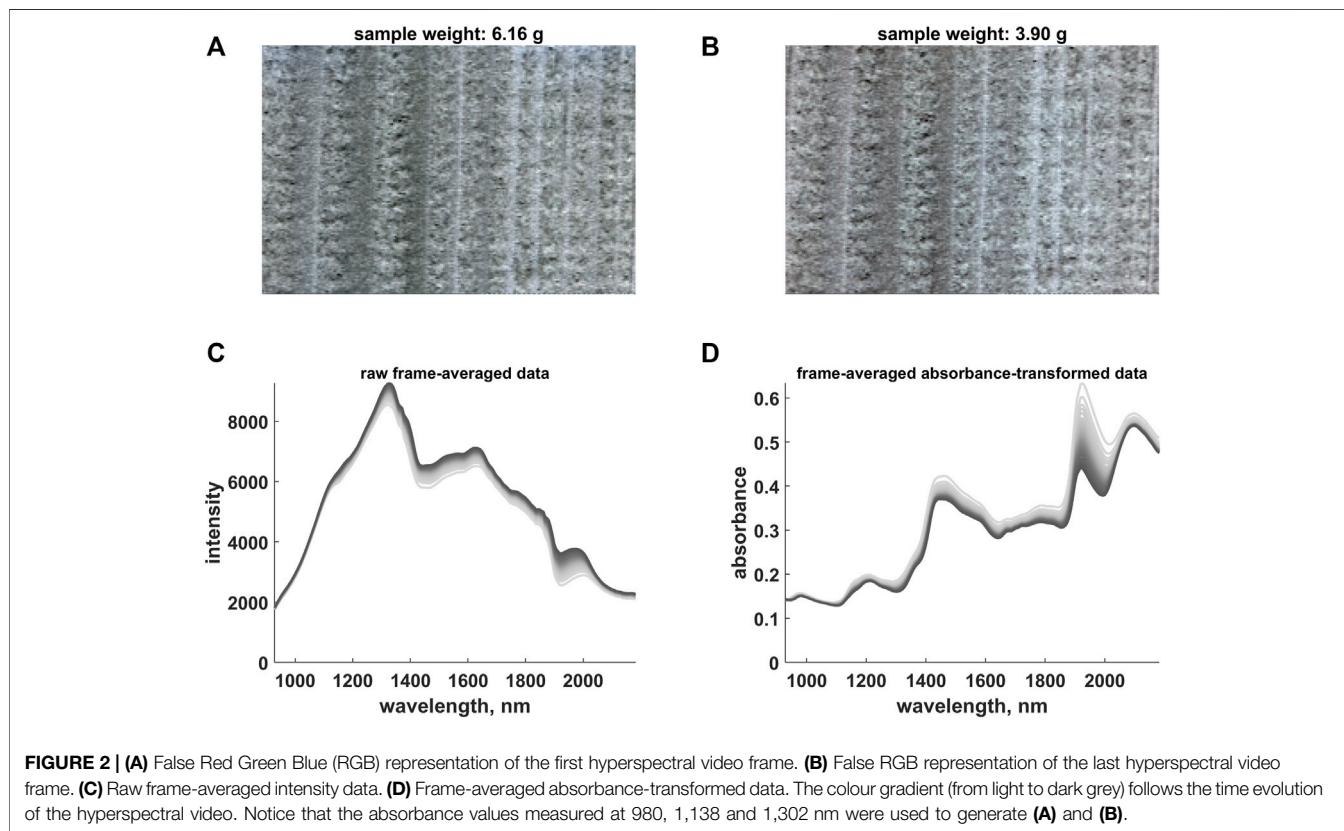
1. the raw data stream, $\mathbf{X}$ (of dimensions, *e.g.*, $N_x N_y K \times J$), divided into a sequence of blocks, say $\mathbf{X}_g$ ($N_g \times J$, $g = 1, 2, \ldots, G$), is submitted to an optional lossless knowledge-based preprocessing stage including a linearisation—which can be conducted by means of approaches like Standard Normal Variate (SNV), Barnes et al. (1989), Multiplicative Scatter Correction (MSC), Martens et al. (1983), Fast Fourier Transform (FFT), Cooley and Tukey (1965), and wavelet decomposition, Walczak (2000)—and a signal-conditioning step;

2. the preprocessed data are projected onto a bilinear subspace already established at the previous point in time as:
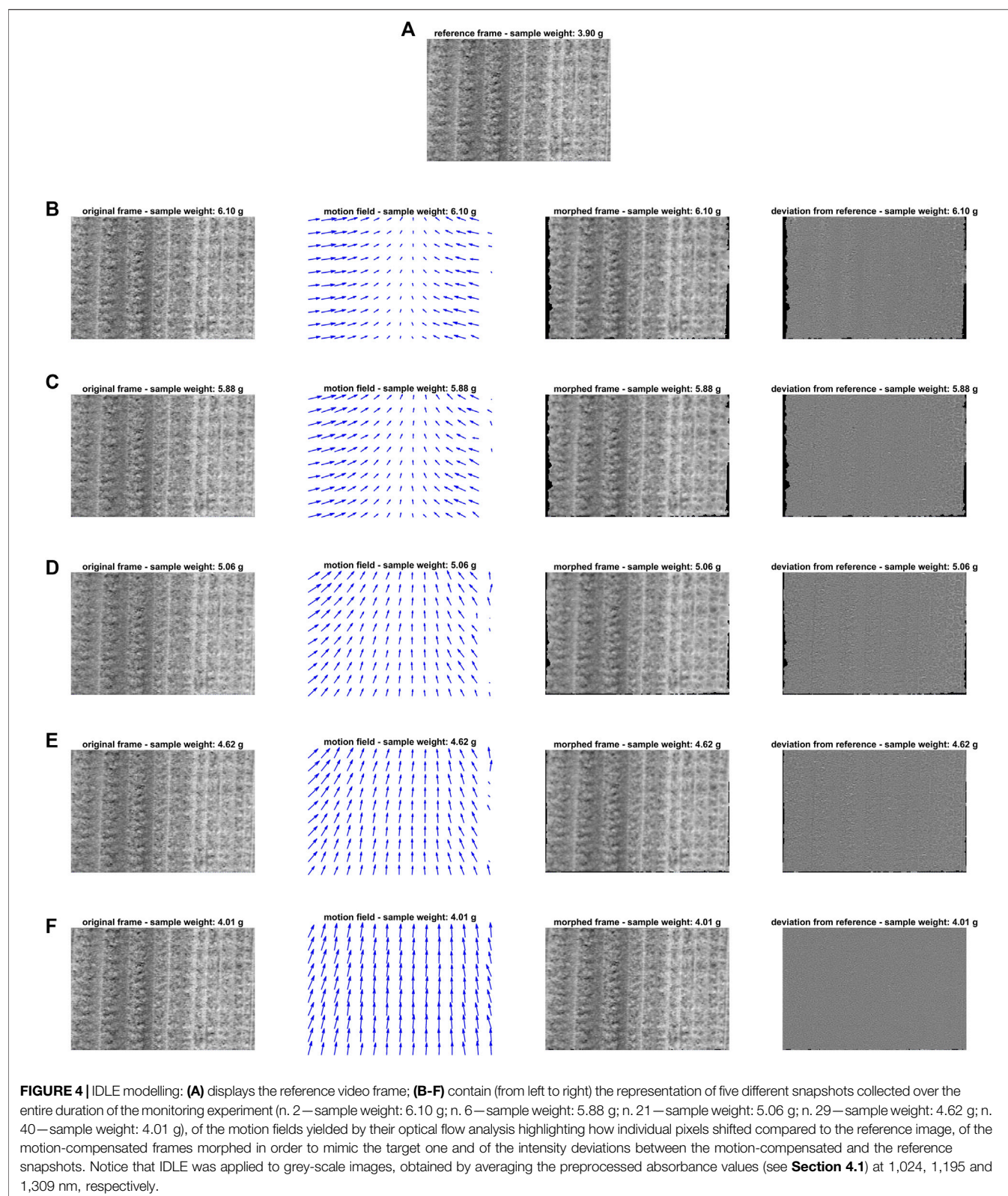
$$\mathbf{X}_g^p = \mathbf{T}_g^p \mathbf{P}^T + \mathbf{E}_g^p \qquad (14)$$

with $\mathbf{T}_g^p$ ($N_g \times A_{\text{OTFP}}$) defining the projection coordinates or scores of all the $N_g$ observations on the basis vectors or components defined by the columns of $\mathbf{P}$ ($J \times A_{\text{OTFP}}$) and $\mathbf{E}_g^p$ ($N_g \times J$) carrying unmodelled residuals, *i.e.*, the fraction of $\mathbf{X}_g^p$ not explained by the model at the chosen rank, $A_{\text{OTFP}} < J$;

3. the projection residuals are thereafter input to a second bilinear modelling stage aimed at detecting new components and isolating outliers. New components are encoded as additional subspace dimensions, whose final number is usually selected based on the total amount of the original data variance that is to be explained, although alternative criteria may also be exploited—Endrizzi et al. (2014); Vitale et al. (2017a); Vitale and Saccenti (2018). In other words, the OTFP algorithm learns to identify and quantify all the systematic types of covariation in the data as they stream, while filtering out random measurement

**FIGURE 2 |** **(A)** False Red Green Blue (RGB) representation of the first hyperspectral video frame. **(B)** False RGB representation of the last hyperspectral video frame. **(C)** Raw frame-averaged intensity data. **(D)** Frame-averaged absorbance-transformed data. The colour gradient (from light to dark grey) follows the time evolution of the hyperspectral video. Notice that the absorbance values measured at 980, 1,138 and 1,302 nm were used to generate **(A)** and **(B)**.



**FIGURE 3 |** Schematic flowchart of the hyperspectral video processing and analysis framework proposed in this article.

**FIGURE 4 |** IDLE modelling: **(A)** displays the reference video frame; **(B-F)** contain (from left to right) the representation of five different snapshots collected over the entire duration of the monitoring experiment (n. 2—sample weight: 6.10 g; n. 6—sample weight: 5.88 g; n. 21—sample weight: 5.06 g; n. 29—sample weight: 4.62 g; n. 40—sample weight: 4.01 g), of the motion fields yielded by their optical flow analysis highlighting how individual pixels shifted compared to the reference image, of the motion-compensated frames morphed in order to mimic the target one and of the intensity deviations between the motion-compensated and the reference snapshots. Notice that IDLE was applied to grey-scale images, obtained by averaging the preprocessed absorbance values (see **Section 4.1**) at 1,024, 1,195 and 1,309 nm, respectively.
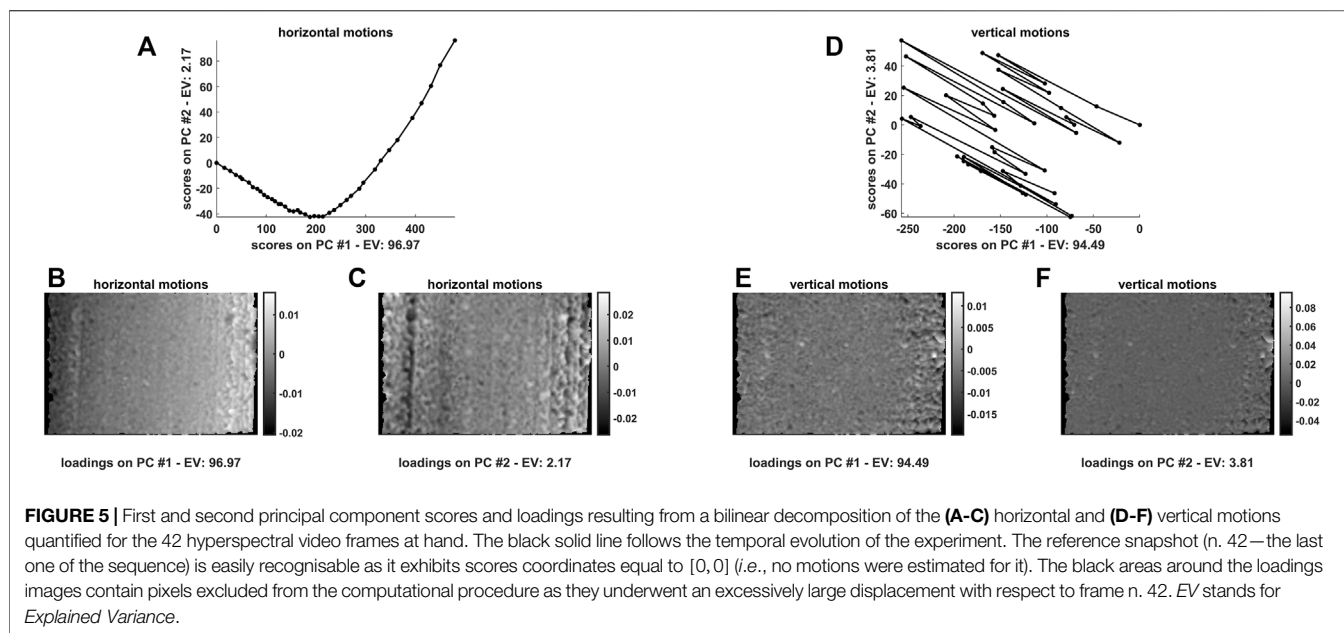
errors and irrelevant outliers (if they do not contribute to the definition of a new pattern of variation);

4. at regular intervals, the OTFP model is refined and updated;

5. pretreatment as well as model parameters (*i.e.*, OTFP scores and loadings) are stored as output. At any time, they can be either used to reconstruct the original data, *e.g.*, for

**FIGURE 5 |** First and second principal component scores and loadings resulting from a bilinear decomposition of the **(A-C)** horizontal and **(D-F)** vertical motions quantified for the 42 hyperspectral video frames at hand. The black solid line follows the temporal evolution of the experiment. The reference snapshot (n. 42—the last one of the sequence) is easily recognisable as it exhibits scores coordinates equal to [0, 0] (*i.e.*, no motions were estimated for it). The black areas around the loadings images contain pixels excluded from the computational procedure as they underwent an excessively large displacement with respect to frame n. 42. *EV* stands for *Explained Variance*.

visualisation, or exploited in their compressed form for efficient storage and transmission, human graphical interpretation and quantification.

A survey of the operational principles of the OTFP is provided in Vitale et al. (2017b).

# 3 DATASET

As model system, a piece of wood of the species Norway Spruce (*Pincea abies*) was submerged in water and soaked for approximately 24 h. Thereafter, it was placed on a digital scale for tracking in real time the variation of its weight and its drying process was monitored by means of a hyperspectral line scan camera (Specim, Oulu, Finland) automatically capturing reflectance images between 930 and 2,200 nm. More specifically, the sample was scanned at regular time intervals, *i.e.*, each time a decrease of around 0.05 g was observed (initial weight: 6.16 g—see **Figure 2A**; final weight: 3.90 g—see **Figure 2B**; total number of hyperspectral images: 42). The sample was illuminated by two halogen lamps positioned on the two sides of the hyperspectral device and never moved during the whole duration of the experiment. A region of interest of $150 \times 225$ pixels was segmented within each frame, which finally resulted in the generation of a four-dimensional dataset of size $150 \times 225 \times 42 \times 200$ (see also **Section 1.3**) and in a memory load of roughly 2.3 GB (double-precision floating-point format).

Although these data were already investigated before—Vitale et al. (2020b)—here, the key role of the linearisation of the instrumental response across space provided by the IDLE approach and its fundamental impact on the assessment and interpretation of the temporal variations of the water signal contributions will be explored.

# 4 RESULTS AND DISCUSSION

A flowchart schematising the general hyperspectral video analysis framework proposed in this work is provided in **Figure 3**.

## 4.1 Spectral Response Linearisation and Frame Greyscale Conversion

In order to compensate the wavelength-dependent variations associated to the light source, the intensity values registered at each $j$th wavelength and at each $n_x \times n_y$-th pixel of the $k$th video frame, $I_{n_x,n_y,k,j}$, were first converted into reflectance units as in :
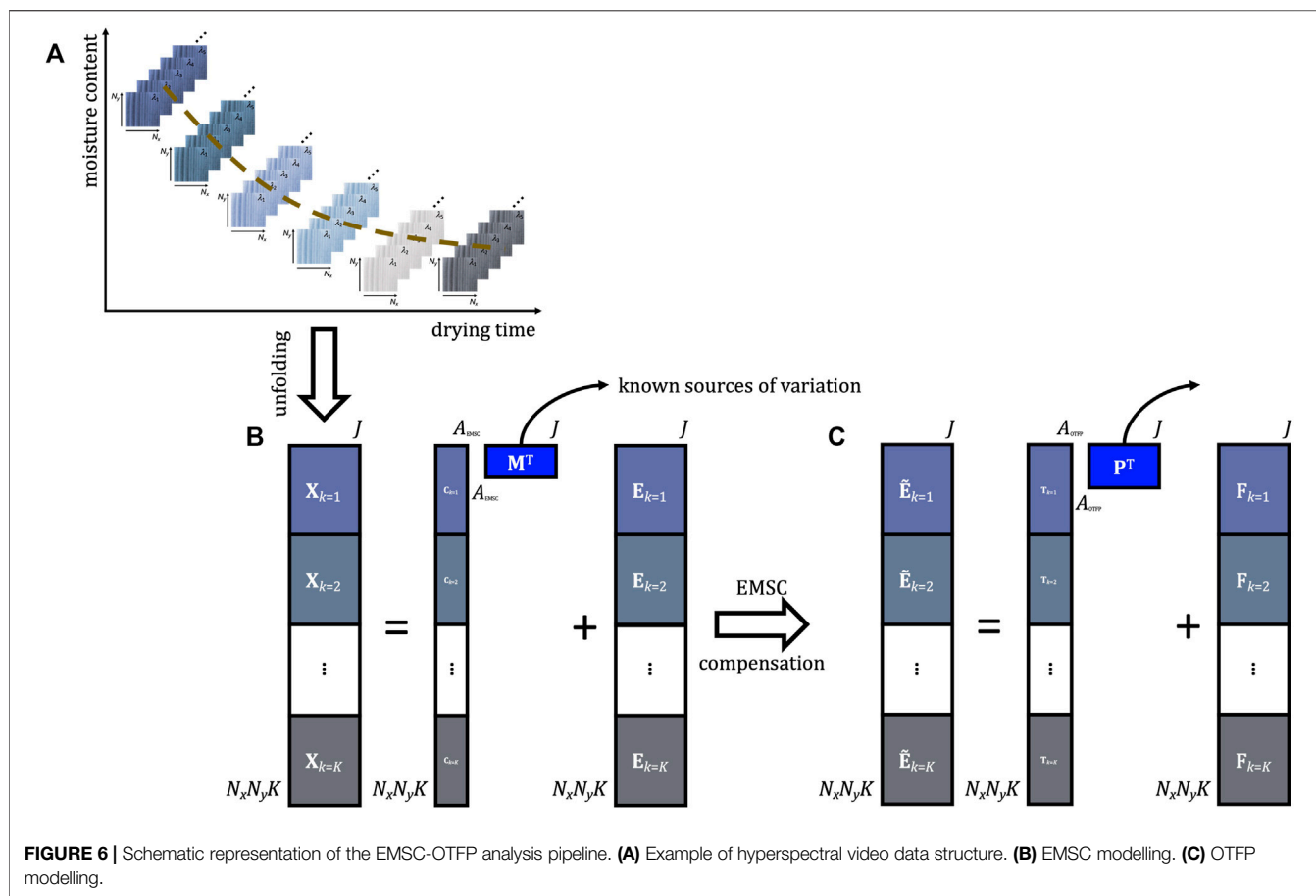
$$R_{n_x,n_y,k,j} = \frac{\left(I_{n_x,n_y,k,j} - I_{n_x,n_y,j,\mathrm{d}}\right)}{\left(I_{n_x,n_y,j,\mathrm{w}} - I_{n_x,n_y,j,\mathrm{d}}\right)} \quad (15)$$

with $I_{n_x,n_y,j,\mathrm{d}}$ and $I_{n_x,n_y,j,\mathrm{w}}$ the intensity recorded at the $j$th wavelength and at $n_x \times n_y$-th pixel for a dark reference and a white reference (a Spectralon sample), respectively. Thereafter, they were transformed into apparent absorbance (that is to say, linearised with respect to the chemical response) according to the following relation:

$$A_{n_x,n_y,k,j} = \log\left(\frac{1}{R_{n_x,n_y,k,j}}\right) = x_{n_x,n_y,k,j} \quad (16)$$

An example of raw and absorbance-converted spectral profiles is provided in **Figure 2C,D**, which highlight the presence of strong baseline variations probably caused by fluctuations in the illumination conditions or in the angular distribution of the reflected light. In order to minimise the bias that such fluctuations (unrelated to sample motions[2]) may induce in the IDLE-based quantification and compensation, an additional

---

[2]For example, those due to water diffusion.

**FIGURE 6 |** Schematic representation of the EMSC-OTFP analysis pipeline. **(A)** Example of hyperspectral video data structure. **(B)** EMSC modelling. **(C)** OTFP modelling.

two-step pretreatment procedure was executed prior to the successive data processing stage:

1. the spectra associated to the pixels of each video frame were pretreated according to an EMSC model similar to the one in **Eq. 8** and encompassing the profiles of two known components: dry wood[3] (reference) and pure water[4]. $\mathbf{W}_{EMSC}$ was set equal to the identity matrix. More specifically, the correction performed for the $n_x \times n_y$-th pixel of the $k$th video frame can be expressed as:

$$\mathbf{x}_{n_x,n_y,k}^{\mathrm{p}} = \frac{\left(\mathbf{x}_{n_x,n_y,k} - a_k\mathbf{1} - d_k\mathbf{f} - g_k\mathbf{f}^2\right)}{b_k} - h_{k,\mathrm{water}}\mathbf{s}_{\mathrm{water}} \quad (17)$$

with $a_k$, $d_k$, $g_k$, $b_k$ and $h_{k,\mathrm{water}}$ being estimated as in **Eq. (9)** from the $k$th frame mean spectrum;

2. at each time point, a grey-scale image, $\mathbf{I}_k$, was then obtained by averaging, for every pixel, the resulting absorbance values at

1,024, 1,195 and 1,309 nm (at these wavelengths, the frame-averaged spectra in **Figure 2D** exhibited the lowest standard deviation). In order to compensate dissimilarities among the intensity cumulative histograms of the various snapshots, these final estimates were ultimately level- and range-adjusted as:

$$\mathbf{I}_k^{\mathrm{p}} = \left(\mathbf{I}_k - \tilde{I}_k\right)\frac{RMS_{\mathrm{ref}}}{RMS_k} + \tilde{I}_{\mathrm{ref}} \quad (18)$$

where $\tilde{I}_k$ and $\tilde{I}_{\mathrm{ref}}$ are the median intensity values within the $k$th and the reference frame (n. 42—sample weight: 3.90 g), respectively, while $RMS_{\mathrm{ref}}$ and $RMS_k$ represent the root-median-squared deviation of the pixel intensities in $\mathbf{I}_{\mathrm{ref}}$ and $\mathbf{I}_k$ from their corresponding median values.

## 4.2 IDLE Modelling

The level- and range-corrected grey-scale images output by the algorithmic procedure outlined in **Section 4.1** were then subjected to IDLE modelling. **Figure 4** summarises the outcomes of the motion estimation-compensation step: **Figure 4A** displays the reference video frame, while, for the sake of illustration, **Figure 4B–F** contain (from left to right) the representation of five other snapshots collected over the entire duration of the monitoring experiment, of the motion fields
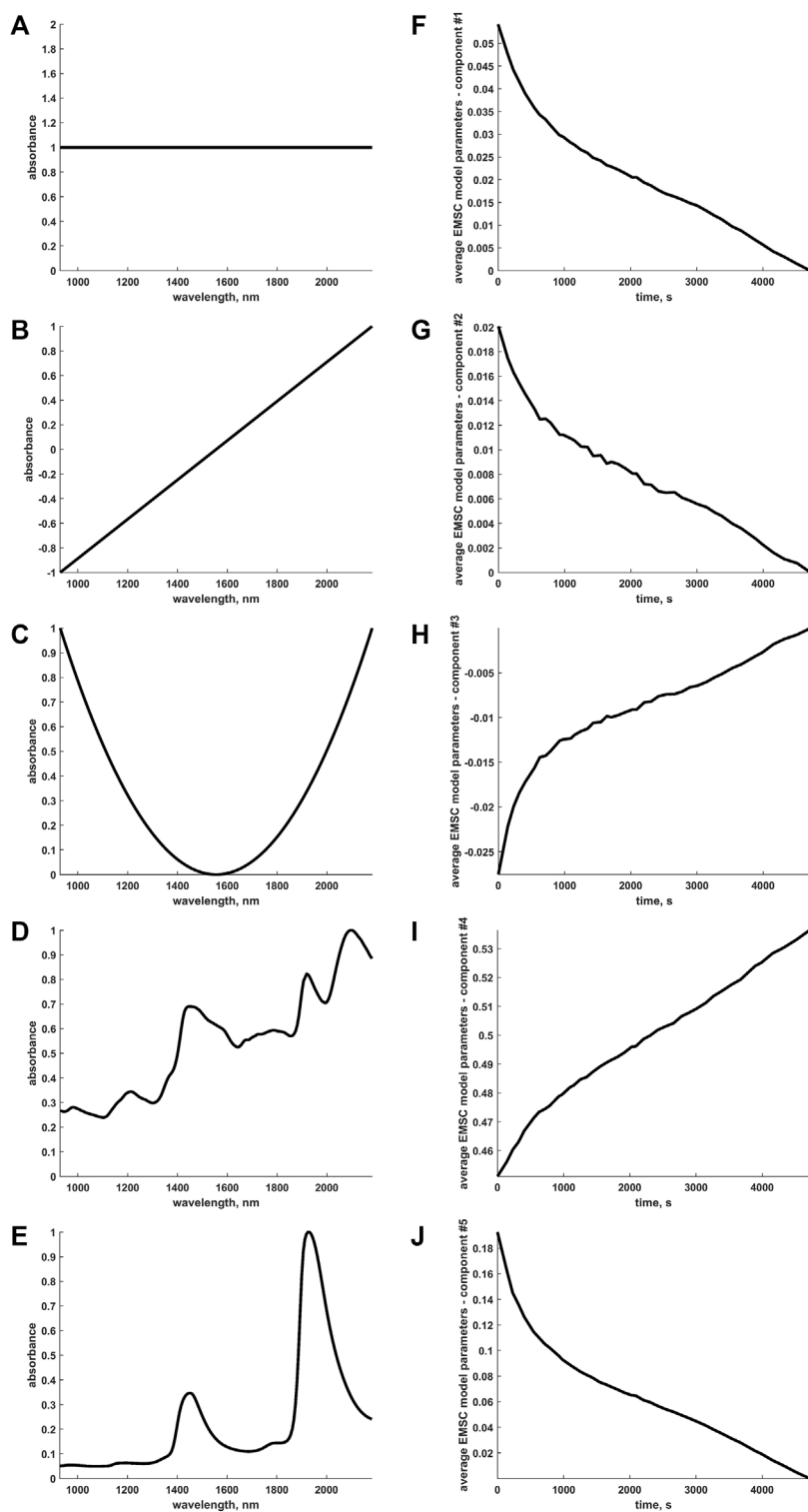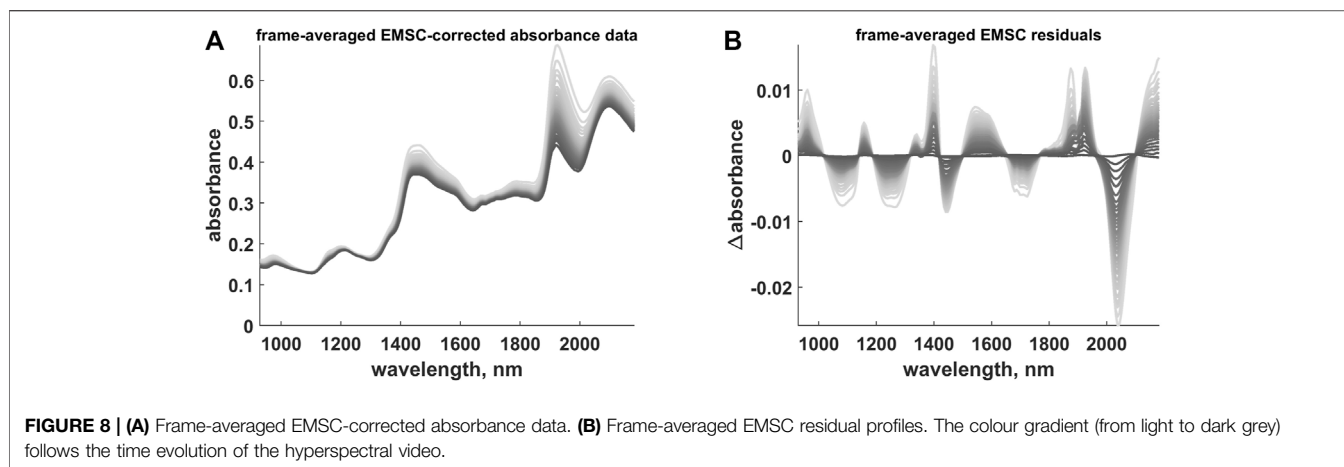
---

[3]Calculated as the average profile of the last video frame.

[4]Measured in reflectance mode by a Nicolet 6700 FT-NIR instrument (Thermo Scientific Inc., Madison, WI, United States) at the same nominal resolution and within the same spectral range as for the hyperspectral video data dealt with in this study and, then, converted into absorbance units.

**FIGURE 7 | (A–E)** Characteristic (absolute max-normalised) spectral profiles submitted to the EMSC computational procedure. The first three represent a typical choice for EMSC modelling as they allow baseline offset, slope and quadratic curvature to be estimated for all the spectra or pixels of the hyperspectral video and compensated before the subsequent On-The-Fly Processing. The last two correspond to the spectra of dry wood and pure water, the two main constituents underlying the specific scene at hand. **(F–J)** Time evolution of the frame-averaged EMSC coefficients (rescaled to compensate the aforementioned absolute max-normalisation) associated to the sources of variation explained by the spectral profiles in **(A–E)**.

**FIGURE 8 | (A)** Frame-averaged EMSC-corrected absorbance data. **(B)** Frame-averaged EMSC residual profiles. The colour gradient (from light to dark grey) follows the time evolution of the hyperspectral video.

yielded by their optical flow analysis highlighting how individual pixels shifted compared to the reference image, of the motion-compensated frames morphed in order to mimic the target one and of the intensity deviations between the motion-compensated and the reference snapshots. As one can clearly see, except for minor edge artefacts, the aforementioned motion fields show how the wood sample horizontally squeezed as it dried and how such horizontal movements significantly decreased at the latest stage of the video recording (*i.e.*, when low amounts of water were present in the pores between wood fibres and compression finally slowed down or stopped). This is also corroborated by the gradual reduction of the number of pixels whose motions could not be properly estimated by IDLE (see the black areas surrounding the motion-compensated frames) because of their relatively large displacement with respect to snapshot n. 42[5]. Notice that these pixels did not undergo EMSC-OTFP processing. Moreover, minimal intensity-deviation-from-target values were observed after image compensation, confirming that wood spatial variations were successfully corrected for.

In order to get additional insights into the nature of such spatial variations along time, the quantified horizontal and vertical motions—retrieved from all the calculated motion fields and concatenated as detailed in Martens (2015)—were analysed by PCA. The resulting temporal scores and spatial loadings are graphed in **Figure 5**.

While vertical shifts appear to follow a random trend (see **Figure 5D**) and might be looked at as mainly due to sideways camera or measurement stage bumps (loadings values are also more or less homogeneously distributed all over the inspected field of view—see **Figures 5A,E,F**) a smoother and more structured evolution was found for the horizontal ones, which further substantiates what stated before about wood squeezing. Horizontal motion scores (see **Figure 5A**) seem to point out the occurrence of a two-phase process during which compression initially proceeds faster and finally decelerates. Horizontal motion loadings along the first principal component (see **Figure 5B**) emphasise the differences between the movements of the pixels of the left and

the right side of the image, while those along the second principal component (see **Figure 5C**) permit to distinguish the distinct behaviour of lateral and central pixels.
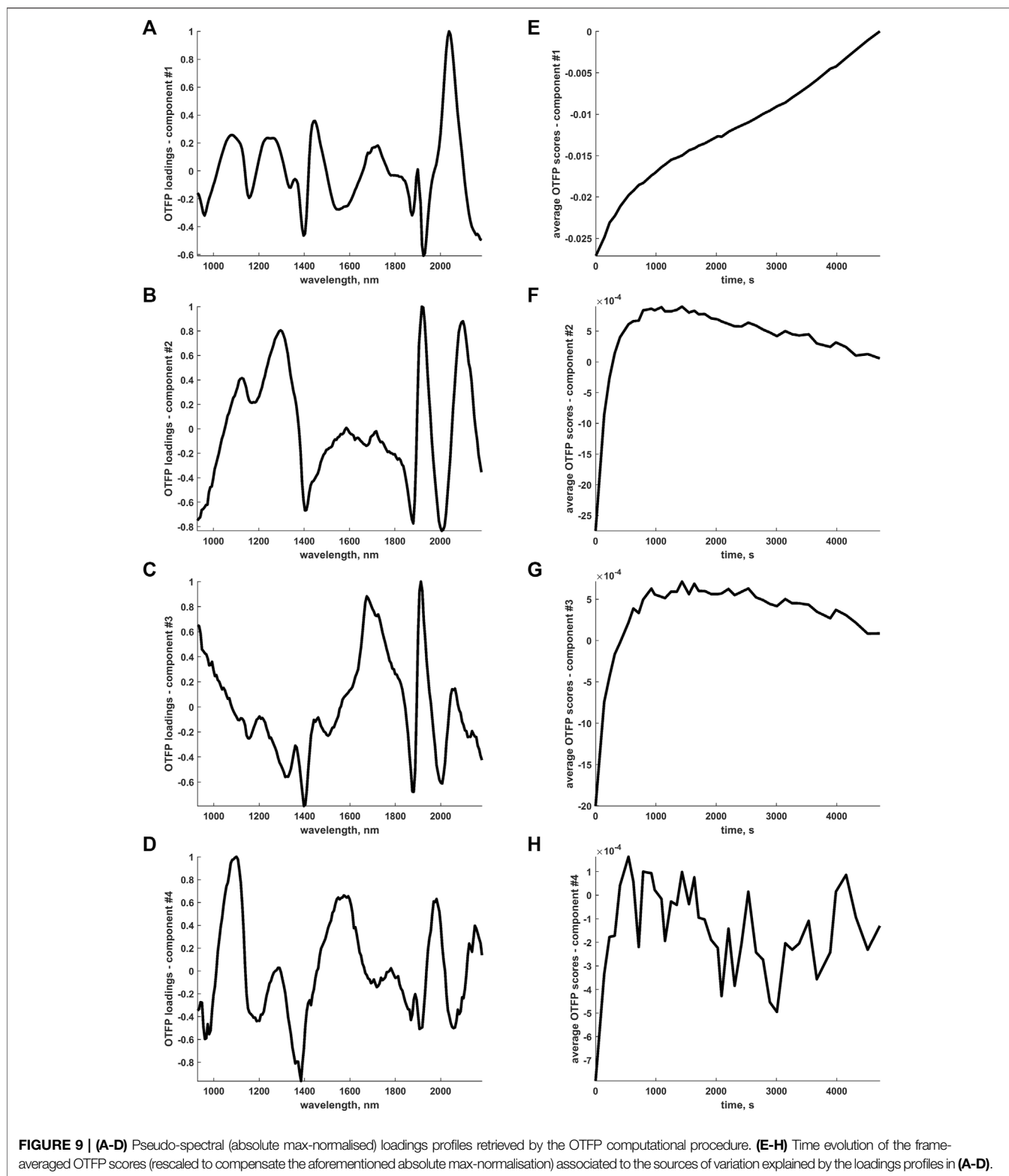
## 4.3 EMSC Modelling

If on the one hand the IDLE approach is capable of quantifying and compensating the movements of a sample observed throughout a hyperspectral video (thus, enhancing the spatial linearity of the instrumental response), on the other hand the combined use of EMSC and the OTFP can enable the identification and retrieval of the most meaningful sources of information from the time series of resulting motion-free hyperspectral images.

The EMSC-OTFP analysis pipeline is schematically outlined in **Figure 6**.

Both EMSC and the OTFP are bilinear modelling techniques that can be utilised in an adaptive- or recursive-like way without requiring entire raw datasets to be kept in memory. The main difference between them regards their respective subspace definition. The matrix **M** (see **Eq. 9** and **Figure 6B**), in fact, is manually constructed by the user based on *a priori* knowledge about the system or the sample under study, which renders EMSC an ideal methodology for extracting and describing expected variation patterns evolving during the progression of a hyperspectral video. On the other hand, **P** (see **Eq. 14** and **Figure 6C**) is automatically learnt by the OTFP algorithm which gradually discovers (in real time) all the sources of systematic variation underlying the data at hand. Consequently, applying sequentially 1) EMSC to the (unfolded) motion-corrected data and 2) the OTFP to the resulting EMSC residuals yields two additive models accounting for both known and unknown phenomena driving the generation mechanism of hyperspectral videos and providing a detailed global overview of the captured dynamic scene.
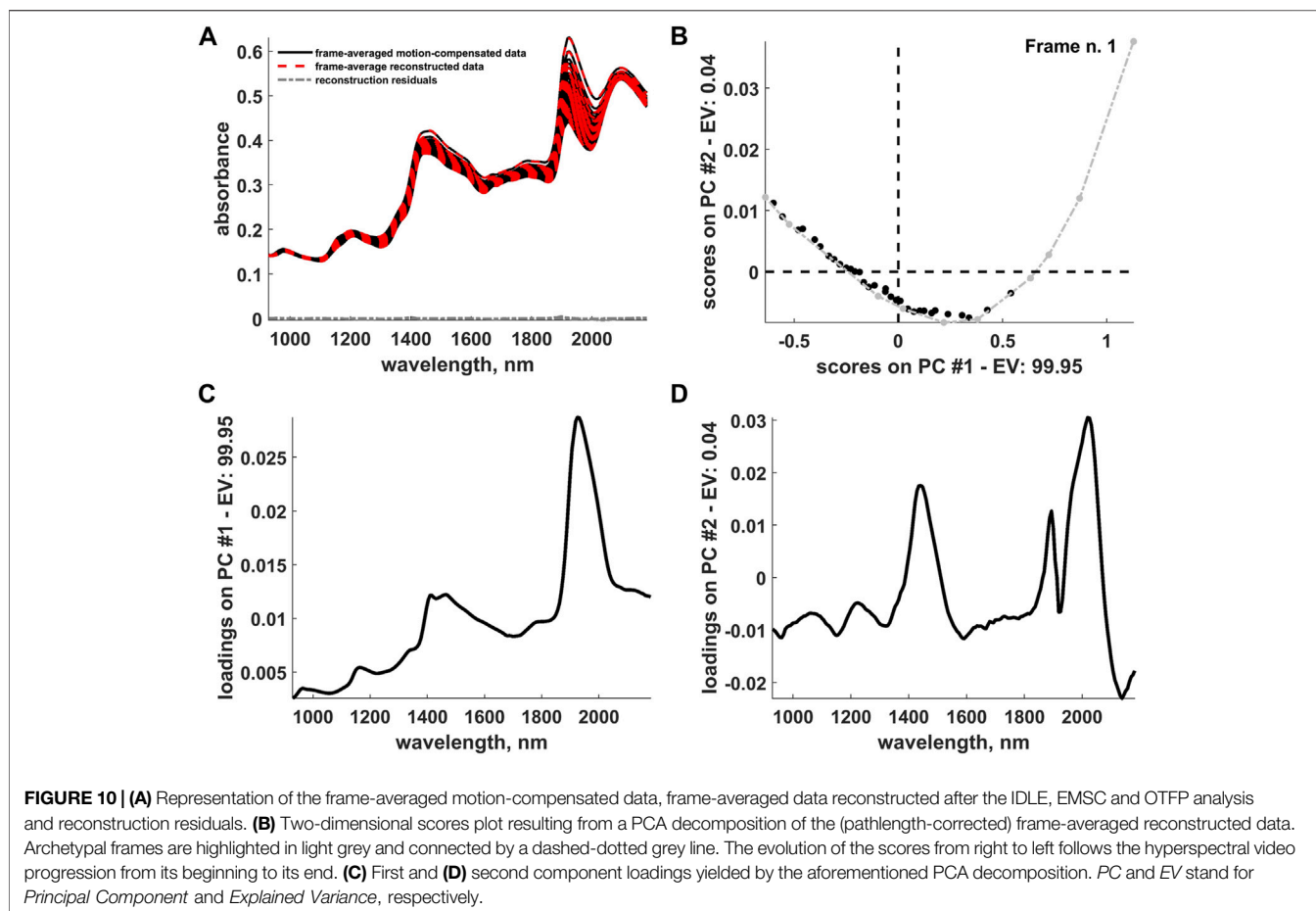
Here, in a first step, the five profiles in **Figure 7A−E** were input to the EMSC algorithmic procedure: as also briefly outlined before, the first three constitute a standard choice for EMSC modelling as they permit to estimate and compensate baseline offset, slope and quadratic curvature for all the pixels of the hyperspectral video before the subsequent application of the OTFP. The last two profiles, instead, correspond to the spectra of dry wood (reference) and pure water, the two major constituents of the specific scene at hand. Representing the

---

[5]In fact, the higher the time difference between frames, the larger the distance that the pixels at the borders of these frames covered due to wood squeezing.

**FIGURE 9 | (A-D)** Pseudo-spectral (absolute max-normalised) loadings profiles retrieved by the OTFP computational procedure. **(E-H)** Time evolution of the frame-averaged OTFP scores (rescaled to compensate the aforementioned absolute max-normalisation) associated to the sources of variation explained by the loadings profiles in **(A-D)**.

time trend of the coefficients yielded for each one of these expected sources of data variability (averaged across all the pixels within every original video frame after motions were compensated, see **Figures 7F–J**) is a simple and immediate way to visualise and assess the

information returned by EMSC and somehow characterise the dynamic evolution of known variability patterns during wood drying. From such graphs, one can easily observe that most of the modelled wood features change quite rapidly within the first stage of

**FIGURE 10 | (A)** Representation of the frame-averaged motion-compensated data, frame-averaged data reconstructed after the IDLE, EMSC and OTFP analysis and reconstruction residuals. **(B)** Two-dimensional scores plot resulting from a PCA decomposition of the (pathlength-corrected) frame-averaged reconstructed data. Archetypal frames are highlighted in light grey and connected by a dashed-dotted grey line. The evolution of the scores from right to left follows the hyperspectral video progression from its beginning to its end. **(C)** First and **(D)** second component loadings yielded by the aforementioned PCA decomposition. *PC* and *EV* stand for *Principal Component* and *Explained Variance*, respectively.

the drying process. This may be due to the residual presence of a thin liquid water film on the surface of the wood sample at the beginning of the hyperspectral monitoring, which could have cloaked its spectral properties. **Figure 7J** also highlights that moisture loss was still proceeding when the experiment was interrupted. Conversely, regarding the wood contribution itself, an approximately constant increasing trend over time was observed. This behaviour accurately reflects the chemical nature of the sample drying which might have been clearly unveiled here because its continuous physical contractions were directly and explicitly accounted for, significantly reducing the spatial complexity of the considered video data. It goes without saying, then, that exploiting simultaneously both spectral and spatial information encoded in hyperspectral videos can significantly enhance the comprehension and understanding of the physico-chemical phenomena behind complex real-world systems.
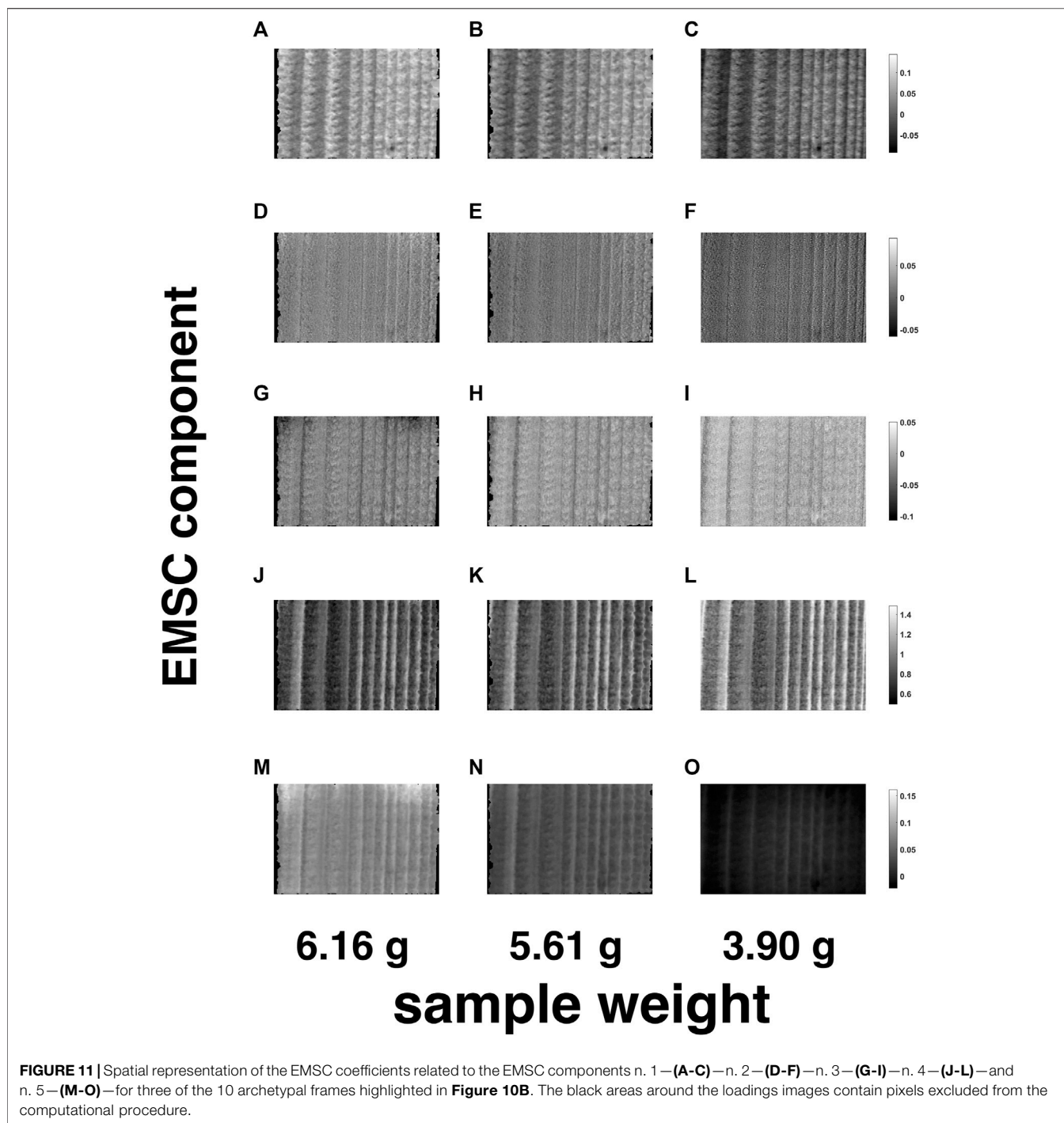
## 4.4 OTFP Modelling

After the EMSC compensation (see **Figure 8A**), the resulting residual profiles (see **Figure 8B**) were submitted to the OTFP for automatically retrieving all the systematic sources of variation left unmodelled by the first data analysis steps[6].

---

[6]4 OTFP components were required to explain around 80% of the EMSC residual variance.

Even if the interpretation of the OTFP output may seem more complicated due to the fact that the OTFP subspace features PCA-like orthogonal bases, smooth and rather well-defined time trends were found for the frame-averaged OTFP coefficients or scores (see **Figure 9E–H**). Such time trends highlight the existence of at least two structured phases in the process of wood drying. Consider, for example, **Figure 9F**: an initial fast transition from negative to positive scores values can be observed followed by a smoother descendant evolution approximately plateauing at around 0. Given also that most of the OTFP loadings profiles in **Figure 9A–D** show large contributions associated to the main water absorption regions, one can reasonably envision the occurrence of more complex phenomena directly related to the thermodynamic state of water itself (*i.e.*, free or bound).
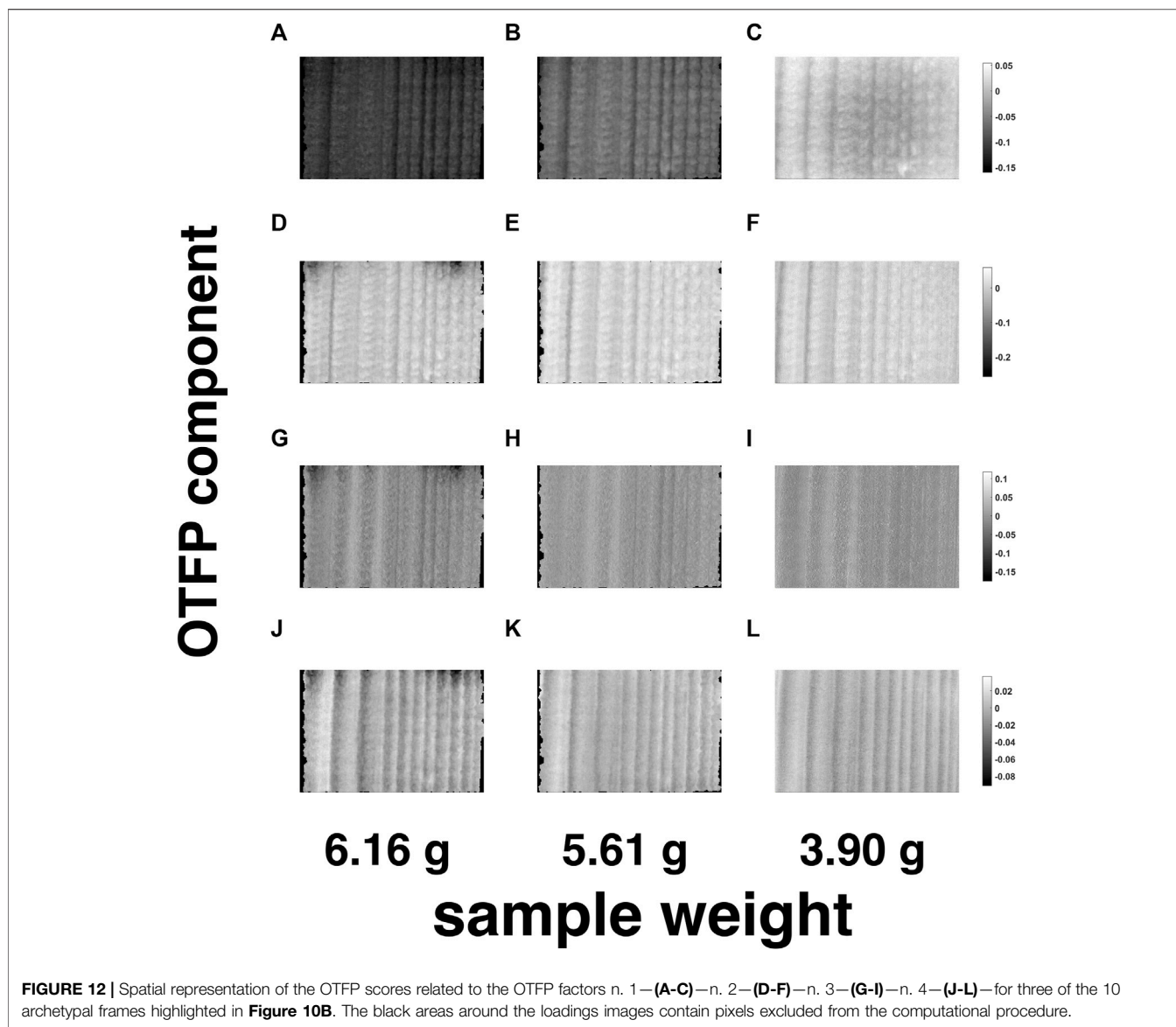
## 4.5 Data Reconstruction and Postprocessing

For a tentative exploration of the thermodynamic phenomena mentioned in **Section 4.4**, the pathlength-corrected absorbance spectra, obtained by reconstructing and averaging the 42 motion-compensated hyperspectral video frames after EMSC and OTFP processing (see **Figure 10A**), were decomposed by standard PCA and graphed in the scores plot in **Figure 10B**. This plot clearly highlights the occurrence of a two-phase transition process

**FIGURE 11 |** Spatial representation of the EMSC coefficients related to the EMSC components n. 1—**(A-C)**—n. 2—**(D-F)**—n. 3—**(G-I)**—n. 4—**(J-L)**—and n. 5—**(M-O)**—for three of the 10 archetypal frames highlighted in **Figure 10B**. The black areas around the loadings images contain pixels excluded from the computational procedure.

during wood drying affecting mainly the water bands of such NIR spectra (see the loadings in **Figures 10C,D**) and characterised by 10 archetypal time instants (see the grey dots in **Figure 10B**)— Ruckebusch et al. (2020). **Figures 11**, **12** provide an illustration of the distribution of the EMSC coefficients and the OTFP scores over the surface of the wood sample at three of these time instants. This representation allows assessing the aforementioned

transition process at a spatial level: overall, the coefficient spatial distribution seems to get smoother as the experiment evolves, which might be explained in the light of the continuous migration/diffusion of water molecules through the pores of the wood specimen (see, *e.g.*, **Figure 12D–F**). However, all these aspects will be investigated in future research also by means of more rational subspace axis rotations—performed, for instance,

**FIGURE 12 |** Spatial representation of the OTFP scores related to the OTFP factors n. 1—**(A-C)**—n. 2—**(D-F)**—n. 3—**(G-I)**—n. 4—**(J-L)**—for three of the 10 archetypal frames highlighted in **Figure 10B**. The black areas around the loadings images contain pixels excluded from the computational procedure.

by varimax, Kaiser (1958), Independent Component Analysis (ICA), Comon (1994); Hyvärinen et al. (2001), or MCR-ALS—aimed at optimising the meaningfulness of the OTFP factors from a physico-chemical perspective.

## 5 CONCLUSION

Hyperspectral videos generate a lot of informative data. However, these data require efficient mathematical modelling for being reliable, understandable and quantitatively interpretable. Here, a general framework by which hyperspectral videos can be analysed was proposed. The three computational steps of this framework result in a compact multi-domain hybrid subspace modelling approach, involving spatial, spectral and temporal parametrisation of

both known and unknown chemical and physical phenomena underlying the studied systems. IDLE permits to characterise and compensate the complex motions that the investigated objects may undergo over the measurement time. EMSC is capable of providing a simple mathematical description of a range of phenomena (and of their temporal evolution) that operators expect or presume *a priori* to be occurring over the duration of the hyperspectral video recording. Finally, the OTFP compresses and summarises all the information related to unknown or unexpected events which may happen during the progression of the data collection. In other words, one can look at the combination of these three different methodologies as an algorithmic extension of how human beings observe reality: the eyes capture spatial changes in the external environment and submit particular signals to the brain that afterwards processes them distinguishing between

what was somehow forecastable in advance (based on past experiences) and what is completely new and unforeseen. In this regard, rather than the individual application of the aforementioned techniques (some of which are already well-established in the field of chemometrics), it is their fusion into a comprehensive algorithmic architecture for the global assessment and interpretation of time-series of high-dimensional hyperspectral images to be innovative and unprecedented.

The sequential IDLE-EMSC-OTFP hybrid framework presented here rests on a combination of targeted and non-targeted data modelling of both known and unknown variation sources. In contrast to classical subspace decomposition strategies (*e.g.*, PCA, PLS, MCR-ALS, NNMF and ICA), it enables the description not only of additive spectral response variations, but also of multiplicative ones (like physical structure effects on the optical pathlength) and hard and soft shape changes (due, for example, to sample repositioning and/or shrinkage).

Moreover, differently from machine learning methods based on Artificial Neural Networks (ANN)—Gasteiger and Zupan (1993)—and Convolutional Neural Networks (CNN)—Gu et al. (2018)—the IDLE-EMSC-OTFP modelling approach yields a strong dimensionality reduction of torrents of input data and results graphically interpretable in their compressed state, revealing how spectral properties, spatial patterns and temporal dynamics are strictly intertwined into unified variation components, whose assessment and interpretation might provide fundamental insights into underlying chemical, physical and instrumental causalities. In the future, relying on a trilinear rather than bilinear OTFP model structure—exploiting, for instance, the principles of Parallel Factor Analysis (PARAFAC), Harshman (1970); Carroll and Chang (1970); Bro (1997)—may enhance this process further.

These conclusions are substantiated and thoroughly corroborated by the outcomes reported in this article. In fact:

1. motion estimation-compensation by spatiotemporal IDLE modelling allowed shrinkage induced by wood drying to be modelled and corrected for, reducing the spatial complexity of the hyperspectral imaging data;
2. EMSC preprocessing permitted a simpler spectral modelling by detecting and disentangling light absorption/light scattering-related variation patterns and their respective evolution over time;
3. the continuous data-driven bilinear subspace decomposition returned by the OTFP enabled the study of the dynamics of the various physical and chemical variations left unmodelled in the stream of hyperspectral residuals after the previous two steps.

In the light of all this and considering its computational efficiency when massive (potentially ever-lasting) flows of multi-channel measurements are handled, the developed approach could have an enormous impact also within the more general context of BIG DATA.

## DATA AVAILABILITY STATEMENT

Data are available under request. Inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

RV and HM conceived and designed the study. IB collected and organised the database. RV and HM performed the statistical data analysis. CR and IB validated both outcomes and conclusions. RV wrote the first draft of the article. RV, CR, IB and HM wrote sections of the paper. All authors contributed to the revision of the manuscript and approved its submitted version.

## REFERENCES

Barnes, R. J., Dhanoa, M. S., and Lister, S. J. (1989). Standard normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* 43, 772–777. doi:10.1366/0003702894202201

Beer, A. (1852). Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys. Chem.* 162, 78–88. doi:10.1002/andp.18521620505

Bouguer, P. (1729). *Essai d'optique sur la gradation de la lumière*. Editor C. A. Jombert. First edn (Paris, France).

Bro, R. (1997). PARAFAC. Tutorial and Applications. *Chemometrics Intell. Lab. Syst.* 38, 149–171. doi:10.1016/s0169-7439(97)00032-4

Carroll, J. D., and Chang, J.-J. (1970). Analysis of Individual Differences in Multidimensional Scaling via an *N*-Way Generalization of "Eckart-Young" Decomposition. *Psychometrika* 35, 283–319. doi:10.1007/bf02310791

Chuvieco, E., and Kasischke, E. (2007). Remote Sensing Information for Fire Management and Fire Effects Assessment. *J. Geophys. Res.* 112, article number G01S90. doi:10.1029/2006jg000230

Comon, P. (1994). Independent Component Analysis, a New Concept? *Signal. Process.* 36, 287–314. doi:10.1016/0165-1684(94)90029-9

Cooley, J. W., and Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Comp.* 19, 297–301. doi:10.1090/s0025-5718-1965-0178586-1

Elmasry, G., Kamruzzaman, M., Sun, D.-W., and Allen, P. (2012). Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: a Review. *Crit. Rev. Food Sci. Nutr.* 52, 999–1023. doi:10.1080/10408398.2010.543495

Endrizzi, I., Gasperi, F., Rødbotten, M., and Næs, T. (2014). Interpretation, Validation and Segmentation of Preference Mapping Models. *Food Qual. Preference* 32, 198–209. doi:10.1016/j.foodqual.2013.10.002

Fischer, C., and Kakoulli, I. (2006). Multispectral and Hyperspectral Imaging Technologies in Conservation: Current Research and Potential Applications. *Stud. Conservation* 51, 3–16. doi:10.1179/sic.2006.51.supplement-1.3

Gasteiger, J., and Zupan, J. (1993). Neural Networks in Chemistry. *Angew. Chem. Int. Ed. Engl.* 32, 503–527. doi:10.1002/anie.199305031

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent Advances in Convolutional Neural Networks. *Pattern Recognition* 77, 354–377. doi:10.1016/j.patcog.2017.10.013

Harshman, R. (1970). Foundations of the PARAFAC Procedure: Models and Conditions for an "explanatory" Multimodal Factor Analysis. *UCLA Working Pap. Phonetics* Vol. 16, 1–84. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.144.5652&rep=rep1&type=pdf

Hay, G. J., Kyle, C., Hemachandran, B., Chen, G., Rahman, M. M., Fung, T. S., et al. (2011). Geospatial Technologies to Improve Urban Energy Efficiency. *Remote Sensing* 3, 1380–1405. doi:10.3390/rs3071380

Horn, B. K. P., and Schunck, B. G. (1993). "Determining Optical Flow": a Retrospective. *Artif. Intelligence* 59, 81–87. doi:10.1016/0004-3702(93)90173-9

Horn, B. K. P., and Schunck, B. G. (1981). Determining Optical Flow. *Artif. Intelligence* 17, 185–203. doi:10.1016/0004-3702(81)90024-2

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* 24, 417–441. doi:10.1037/h0071325

Hugelier, S., Vitale, R., and Ruckebusch, C. (2020). *Image Processing in Chemometrics. Comprehensive Chemometrics*. Second edn., Vol. 4. Amsterdam, Netherlands: Elsevier, B.V., 411–436. doi:10.1016/b978-0-12-409547-2.14597-4:

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. First edn. Hoboken: United States of America: John Wiley & Sons, Ltd.

Kaiser, H. F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika* 23, 187–200. doi:10.1007/bf02289233

Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., and Abbas, A. (2018). Modern Trends in Hyperspectral Image Analysis: a Review. *IEEE Access* 6, 14118–14129. doi:10.1109/access.2018.2812999

Kubelka, P., and Munk, F. (1931). An Article on Optics of Paint Layers. *Z. Tech. Phys.* 12, 593–601. Available at: https://www.graphics.cornell.edu/westin/pubs/kubelka.pdf

Lambert, J. (1760). in *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Editor C. P. Detleffsen. First edn (Augsburg, Gerrmany.

Lawton, W. H., and Sylvestre, E. A. (1971). Self Modeling Curve Resolution. *Technometrics* 13, 617–633. doi:10.1080/00401706.1971.10488823

Lu, G., and Fei, B. (2014). Medical Hyperspectral Imaging: a Review. *J. Biomed. Opt.* 19, article number 010901. doi:10.1117/1.jbo.19.1.010901

Makeig, S., Enghoff, S., Jung, T., and Sejnowski, T. (2000). "Moving-window ICA Decomposition of EEG Data Reveals Event-Related Changes in Oscillatory Brain Activity," in Proceedings of the Second International Workshop on Independent Component Analysis and Signal Separation, 627–632.

Martens, H. (1979). Factor Analysis of Chemical Mixtures. *Analytica Chim. Acta* 112, 423–442. doi:10.1016/s0003-2670(01)85040-6

Martens, H., Jensen, S., and Geladi, P. (1983). "Multivariate Linearity Transformation for Near-Infrared Reflectance Spectrometry," in Proceedings of the Nordic Symposium on Applied Statistics, 205–234.

Martens, H., and Næs, T. (1989). *Multivariate Calibration*. First edn. Hoboken: United States of America: John Wiley & Sons, Ltd.

Martens, H., Nielsen, J. P., and Engelsen, S. B. (2003). Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. *Anal. Chem.* 75, 394–404. doi:10.1021/ac020194w

Martens, H. (2015). Quantitative Big Data: where Chemometrics Can Contribute. *J. Chemometrics* 29, 563–581. doi:10.1002/cem.2740

Martens, H., and Stark, E. (1991). Extended Multiplicative Signal Correction and Spectral Interference Subtraction: New Preprocessing Methods for Near Infrared Spectroscopy. *J. Pharm. Biomed. Anal.* 9, 625–635. doi:10.1016/0731-7085(91)80188-f

Matikainen, L., and Karila, K. (2011). Segment-Based Land Cover Mapping of a Suburban Area-Comparison of High-Resolution Remotely Sensed Datasets Using Classification Trees and Test Field Points. *Remote Sensing* 3, 1777–1804. doi:10.3390/rs3081777

McMillen, J. (1964). Wood Drying-Techniques and Economics. Approved Technical Article, Food Products Laboratory, Forest Service, U.S. Department of Agriculture. Available at: https://www.fpl.fs.fed.us/documnts/pdf1964/mcmil64a.pdf

Pearson, K. (1901). LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Phil. Mag. J. Sci.* 2, 559–572. doi:10.1080/14786440109462720

Ruckebusch, C., Vitale, R., Ghaffari, M., Hugelier, S., and Omidikia, N. (2020). Perspective on Essential Information in Multivariate Curve Resolution. *Trac Trends Anal. Chem.* 132, article number 116044. doi:10.1016/j.trac.2020.116044

Silva, C. S., Pimentel, M. F., Amigo, J. M., Honorato, R. S., and Pasquini, C. (2017). Detecting Semen Stains on Fabrics Using Near Infrared Hyperspectral Images and Multivariate Models. *Trac Trends Anal. Chem.* 95, 23–35. doi:10.1016/j.trac.2017.07.026

Stefansson, P., Fortuna, J., Rahmati, H., Burud, I., Konevskikh, T., and Martens, H. (2020). *Hyperspectral Time Series Analysis: Hyperspectral Image Data Streams Interpreted by Modeling Known and Unknown Variations*. First edn., Vol. 32. Amsterdam, Netherlands: Elsevier, B.V., 305–331. doi:10.1016/b978-0-444-63977-6.00014-6:

Tauler, R., Smilde, A., and Kowalski, B. (1995). Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution. *J. Chemometrics* 9, 31–58. doi:10.1002/cem.1180090105

Vitale, R., Hugelier, S., Cevoli, D., and Ruckebusch, C. (2020a). A Spatial Constraint to Model and Extract Texture Components in Multivariate Curve Resolution of Near-Infrared Hyperspectral Images. *Analytica Chim. Acta* 1095, 30–37. doi:10.1016/j.aca.2019.10.028

Vitale, R., and Saccenti, E. (2018). Comparison of Dimensionality Assessment Methods in Principal Component Analysis Based on Permutation Tests. *Chemometrics Intell. Lab. Syst.* 181, 79–94. doi:10.1016/j.chemolab.2018.08.008

Vitale, R., Stefansson, P., Marini, F., Ruckebusch, C., Burud, I., and Martens, H. (2020b). *Fast Analysis, Processing and Modeling of Hyperspectral Videos: Challenges and Possible Solutions. Comprehensive Chemometrics*. Second edn., Vol. 4. Amsterdam, Netherlands: Elsevier, B.V., 395–409. doi:10.1016/b978-0-12-409547-2.14605-0:

Vitale, R., Westerhuis, J., Næs, T., Smilde, A., de Noord, O., and Ferrer, A. (2017a). Selecting the Number of Factors in Principal Component Analysis by Permutation Testing - Numerical and Practical Aspects. *J. Chemometr* 31, article number e2937. doi:10.1002/cem.2937

Vitale, R., Zhyrova, A., Fortuna, J. F., de Noord, O. E., Ferrer, A., and Martens, H. (2017b). On-The-Fly Processing of Continuous High-Dimensional Data Streams. *Chemometrics Intell. Lab. Syst.* 161, 118–129. doi:10.1016/j.chemolab.2016.11.003

Walczak, B. (2000). *Wavelets in Chemistry, Data Handling in Science and Technology*. First edn. Amsterdam, Netherlands: Elsevier, B.V.

Wang, X., Kruger, U., and Irwin, G. W. (2005). Process Monitoring Approach Using Fast Moving Window PCA. *Ind. Eng. Chem. Res.* 44, 5691–5702. doi:10.1021/ie048873f

Westad, F., and Martens, H. (1999). Shift and Intensity Modeling in Spectroscopy-General Concept and Applications. *Chemometrics Intell. Lab. Syst.* 45, 361–370. doi:10.1016/s0169-7439(98)00144-0

Wold, S., Martens, H., and Wold, H. (1983). *Matrix Pencils. Lecture Notes in Mathematics*. in *Chap. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method*. First edn., Vol. 973. Berlin/Heidelberg, Germany: Springer-Verlag, 286–293. doi:10.1007/bfb0062108