Systematic Assessment of MODTRAN Emulators for Atmospheric Correction

Jorge Vicent-Servera, Juan Pablo Rivera-Caicedo, Jochem Verrelst, Jordi Muñoz-Marí, Neus Sabater, Béatrice Berthelot, Gustau Camps-Valls, Fellow, IEEE, and José Moreno, Senior Member, IEEE

Abstract—Atmospheric radiative transfer models (RTMs) simulate the light propagation in the Earth’s atmosphere. With the evolution of RTMs, their increase in complexity makes them impractical in routine processing such as atmospheric correction. To overcome their computational burden, standard practice is to interpolate a multidimensional lookup table (LUT) of precomputed, simulated outputs. However, accurate interpolation relies on large LUTs, which still implies large computation times for their generation and interpolation. In recent years, emulation has been proposed as an alternative to LUT interpolation. Emulation approximates the RTM outputs by a statistical regression model trained with a low number of RTM runs. However, a concern is whether the emulator reaches sufficient accuracy for atmospheric correction. Therefore, we have performed a systematic assessment of key aspects that impact the precision of emulating MODTRAN: 1) regression algorithm; 2) training database size; 3) dimensionality reduction (DR) method and a number of components; and 4) spectral resolution. The Gaussian processes regression (GPR) was found the most accurate emulator. The principal component analysis remains a robust DR method and among the training databases used, 1000 samples covering a broad range of atmospheric conditions, GPR emulators can reconstruct the simulated spectral data with relative errors below 1% for the 95th percentile. These emulators reduce the processing time from days to minutes, preserving sufficient accuracy for atmospheric correction and providing model uncertainties and derivatives. We provide a set of guidelines and tools to design and generate accurate emulators for satellite data processing applications.

Index Terms—Atmospheric correction, emulation, hyperspectral, MODTRAN, radiative transfer.

I. INTRODUCTION

ATMOSPHERIC correction is one of the main algorithms when processing optical satellite data [1]. Its main task is the conversion of the top-of-atmosphere (TOA) radiance signal measured by a satellite instrument into surface reflectance. After compensation of the atmospheric effects, the derived surface reflectance data can be used to retrieve geophysical properties for applications, such as vegetation monitoring [2] or water quality [3]. In general, atmospheric correction algorithms can be categorized into two main families: 1) semi-empirical and 2) physically based methods. The first category refers to image-based methods that aim at normalizing the TOA radiance data through semi-empirical approaches [4]. The second category refers to methods that make use of radiative transfer models (RTM) to derive the atmospheric composition and to decouple the surface reflectance from atmospheric transfer functions (i.e., path radiance, transmissions, and spherical albedo) [5]–[7]. Because of such a rigorous approach, the latter methods are more accurate, though slower, than the first category. Atmospheric RTMs are computer models that describe the physical processes of scattering, absorption, and emission of the electromagnetic radiation within the Earth’s atmosphere [8]. They are widely used in the atmospheric correction and other Earth observation applications, such as scene generation [9], atmospheric chemistry [10], and numerical weather prediction [11]. Over time, these models have increased their realism from simple semi-empirical models [12] toward advanced physically based models, such as MODTRAN [13], MOMO [14], and libRadtran [15]. This evolution has led to an increase in computational time and memory requirements to run the model, making RTMs impractical in routine data processing chains, such as atmospheric correction. To overcome these limitations, the common approach is to interpolate lookup tables (LUTs) of precalculated RTM simulations [16]. However, large LUTs are still needed to achieve sufficient accuracies, imposing long computation times in case of slow RTMs [17]. In addition, advanced interpolation techniques are memory intensive, and thus, limited to low-dimensional spaces [18], [19].

In this context, RTM emulators were proposed as an accurate and fast alternative to LUT interpolation [20], [21].

Manuscript received November 9, 2020; revised March 30, 2021; accepted April 2, 2021. This work was supported by the FLEX L2 End-to-End Simulator Development and Mission Performance Assessment Project, European Space Agency (ESA), under Contract 4000119707/17/NL/MP. The work of Jochem Verrelst was supported by the European Research Council (ERC) through the ERC-2017-STG SENTIFLEX Project under Grant 755617. The work of Neus Sabater was supported by the Academy of Finland FluoSYN Project under Grant 324091. The work of Gustau Camps-Valls was supported by the European Space Agency (ESA), under Contract 4000119707/17/NL/MP. The work of Jordi Muñoz-Marí was supported by the ERC through the ERC-CoG-2014 SEDAL Project under Grant 647423. The work of Jorge Vicent-Servera is with the Departamento of Secretaría de investigación y postgrado, Consejo Nacional de Ciencia y Tecnología, Autonomous University of Nayarit, Tepic 63173, Mexico (e-mail: jpvicenta@conacyt.mx).

Jochem Verrelst, Jordi Muñoz-Marí, Gustau Camps-Valls, and José Moreno are with the Image Processing Laboratory, University of Valencia, 46100 Valencia, Spain (e-mail: jochem.verrelst@uv.es; jordi.munoz@uv.es; gustau.camps@uv.es; jose.moreno@uv.es).

Neus Sabater is with the Finnish Meteorological Institute, 00560 Helsinki, Finland (e-mail: neus.sabater@fmi.fi).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TGRS.2021.3071376.

Digital Object Identifier 10.1109/TGRS.2021.3071376

An emulator is a surrogate statistical learning model that approximates the original deterministic model at a fraction of the original running time [22], [23]. Therefore, emulation and LUT interpolation serve essentially the same purpose: to approximate the original atmospheric RTM outputs. The idea of emulating atmospheric RTMs has already been successfully tested to global sensitivity analysis [24], [25]. In these works, the emulators were designed so that their accuracy was sufficient to study the dependencies of the RTM outputs on the main input variables through a statistical global sensitivity analysis. However, when it comes to applying an emulator in a data processing scheme, a critical aspect is a precise approximation of the original RTM. In fact, various factors determine the accuracy and performance of the emulator [26]: 1) used machine learning method; 2) training database size; and 3) applied dimensionality reduction (DR) strategy. These three aspects must be systematically assessed in order to develop an emulator that is both fast and accurate to be used for atmospheric correction as an alternative over LUT interpolation.

Therefore, the main goal of this work is to perform a systematic assessment of various emulator configurations in terms of computation time and accuracy to reconstruct MODTRAN-based atmospheric transfer functions for a wide range of atmospheric conditions. This work focuses on the following two aspects. The first aspect refers to the identification of the most suitable statistical learning method by evaluating the role of: 1) regression algorithm and DR strategy. The second aspect aims at further optimizing the emulator by studying the role of: 2) a number of components, as a function of the spectral resolution and 3) the training database size. As a proof-of-concept, the best performing MODTRAN emulator will be applied for the atmospheric correction of real Sentinel-3 and synthetic TOA radiance from the future Fluorescence Explorer (FLEX) mission, European Space Agency (ESA), Paris, France. We will close this article with a discussion on potential applications of atmospheric RTM emulators.

II. EMULATION THEORY

Emulation is a regression model that approximates model simulations from the statistical learning of a training data set [23]. Similar to LUT interpolation, an emulator uses a database of prestored model simulations (i.e., training samples) to infer the output values of a computationally expensive code for an unseen input configuration. Therefore, an emulator is an advanced regression model as typically done for retrieval applications but then in reversed order [27], [28]. Emulation offers a practical solution for routine applications to efficiently approximate computer codes of high computational burden, which is the case of most atmospheric RTMs. This technique has been successfully applied for the last few decades in the climate and environmental modeling communities [29]–[31] and to approximate atmospheric RTMs [24], [32]. The emulation technique is possible thanks to the development of adaptive and flexible regression algorithms [33], with successful examples, such as Gaussian processes regression (GPR) and neural networks (NNs) [20], [34]. An important aspect to take into account when replacing an RTM with an emulator is that one should ensure the same properties of the model; the emulator should be mathematically tractable and permit the calculation of the Jacobians, uncertainty quantification and propagation, and sensitivity analysis. As LUT-based methods, emulators also keep these properties by computing the Jacobians numerically and, for kernel methods, also analytically [35]. This allows error quantification and propagation as well sensitivity analysis from an emulator, features that are not commonly implemented within an RTM code [11].

The emulation technique is, however, challenged when it comes to predict the multiple outputs produced by an RTM code and reconstruct a full-spectral simulation. Indeed, a full 400–2500-nm spectral data consists of over 10 000 bands when simulated at 1 cm$^{-1}$ resolution. This is an additional difficulty compared with classical emulators since only a few regression algorithms are capable of generating multiple outputs [36]. Moreover, it can take a considerable time to train such a complex statistical multi-output model, especially when using NN [37]. A workaround solution is to take advantage of the Hughes phenomenon [38] and the large degree of co-linearity in spectroscopic data. Accordingly, a DR method can transform the spectral data into a lower dimensional space, in which the number of components is a fraction of the original amount of spectral bands. With the application of a DR method, the multi-output problem is greatly reduced to a number of components that retain the spectral information from the original data. By first applying a DR method, the spectroscopic data is reduced to a given number of components. Then, by looping over each component, multiple models can be trained by single-output regression algorithms. Afterward, the spectral signal can again be reconstructed. Although this iterative process takes some computational time, it goes much faster than training an emulator for every single-spectral channel and makes the problem better conditioned [20], [31], [37]. Moreover, the application of a DR method in the emulation processing allows converting any single-output regression algorithm into a multi-output algorithm. However, in practice, only statistical regression algorithms are sufficiently adaptive to enable approximating the behavior of an RTM.

A. Statistical Regression Algorithms

In order to approximate an RTM (e.g., MODTRAN) through emulation, a statistical regression algorithm must be trained using a database of precalculated RTM simulations [23]. The training data set should ideally cover the multidimensional input space so that the statistical regression uncovers the underlying correlations between model inputs and outputs. One strategy is using a space-filling pseudorandom sampling algorithm, such as Latin hypercube sampling (LHS) [39]. An alternative is to use active learning methods to set the samples in optimum locations of the input space [40]. From our previous literature review [24], [37], we evaluated the following algorithms to function as emulators, belonging to the three most standard families of regression algorithms: 1) the standard and efficient random forest (RF); 2) kernel ridge regression (KRR) and GPR as paradigmatic examples of
kernel methods, and 3) the standard feed-forward multilayer perceptron from the family of artificial NNs.

Decision trees are predictive models based on a set of hierarchical connected nodes, each of them representing a linear decision based on a specific input feature. To overcome the difficulty of classical decision trees to cope with strong nonlinear input–output relationships, a combination of decision trees (RFs) can improve results [41]. We tuned an RF trying with different configurations involving the number of trees in the forest (from 100 to 1000), minimum leaf size (from one to five samples), and a number of variables to select randomly in each split of the trees (from 1/3 of the number of variables, as originally proposed in [41], to all variables).

Artificial NNs are layered circuits of artificial neurons connected via weights (links) representing excitation or inhibition states [42]. Designing and training a NN implies selecting the number of hidden layers and nodes per layer, the shape of the nonlinearity connection, the learning rate, the regularization parameters to avoid overfitting, and initializing the weights. Both the training algorithm and the loss function have an impact on the performance of an NN. Here, we have used the standard multilayer perceptron, which is a fully connected network. In order to simplify its design, the NNs in this work are consisting of only one hidden layer of neurons. The NN structure is optimized using the Levenberg–Marquardt learning algorithm with a squared loss function.

In machine learning, kernel methods use kernel function to quantify similarities (or distances) between input samples of a data set [33], [43]. The similarity is computed by a linear dot product in a higher dimensional feature space, yet without ever computing the data location in the feature space. The following two methods are tested here: 1) KRR [44] and 2) GPR, as a probabilistic version of KRR [45]. For all kernel methods, we used a standard radial basis kernel function.

The above-mentioned regression methods are popular in various application domains thanks to processing times, accuracy, and robustness to handle overfitting. A detailed description of these methods is given in [28] and the simpleR package [46].

B. Dimensionality Reduction Algorithms

As previously introduced, an emulator should combine a regression algorithm with a DR method in order to reconstruct the spectral output generated by an RTM. The regression algorithm first predicts the components of the RTM data in a lower dimensional space, and the inverse of the DR transformation is then applied to reconstruct output spectra at expenses of some loss in accuracy. The principal component analysis (PCA) [47] is a DR method that is widely used to reduce the dimensionality of large data sets. PCA has shown its suitability to reconstruct satellite data and to speed-up atmospheric RTMs [48]–[52]. This method looks for correlations between the spectral data in order to identify a set of orthogonal linearly uncorrelated directions that maximize the variance of the projections. The eigenvectors and eigenvalues of this transformation are estimated from the covariance matrix of the spectral data set X. The eigenvectors matrix, U, is then used as a projection matrix that allows obtaining the so-called X-scores, simply by W = UX. As U is an orthogonal matrix the reconstruction of X given the scores is obtained by X = U⊤W. Hence, the spectra are reconstructed. The DR by PCA is achieved by selecting a number of components p that is much lower than the number of original spectral channels B. Typically, one takes enough components to ensure that at least 99.99% of variance is retained. While this variance level is quite standard, it might also retain noise in the real data.

An alternative DR method that allows reconstruction of the spectral output is the partial least squares (PLS) [53]. The PLS method looks for projections that maximize the covariance and the correlation between the RTM output spectra and its input variables. The possibility to combine multiple regression algorithms with either PCA or PLS allows construction of different emulators. The next step is to assess their performance in order to consolidate an optimized emulator design for atmospheric RTMs and atmospheric correction applications.

III. MATERIALS AND METHODS

The systematic evaluation of emulator configuration was carried out with the in-house developed automated RTMs operator (ARTMO) framework [54]. ARTMO is a scientific software package developed in MATLAB that provides tools and toolboxes for running a suite of leaf, canopy, and atmosphere RTMs and for postprocessing applications, such as emulation. ARTMO has been made compatible with the atmospheric LUT generator (ALG) toolbox [25], a standalone software tool that allows generating LUTs based on a suite of atmospheric RTM, such as MODTRAN, 6SW, or libRadtran. ARTMO and ALG are freely downloadable at www.artmotoolbox.com.

A. Atmospheric Lookup Table Generator

ALG is a MATLAB-compiled software tool that facilitates the generation of LUTs for a wide range of atmospheric RTMs [25]. This tool provides a graphical user interface with which users can configure and execute atmospheric RTM codes for simulations in the optical domain. ALG automatically processes and standardizes all the RTM input and output data into the final LUT file. These LUTs store the so-called atmospheric transfer functions. These spectral outputs permit uncoupling the atmospheric absorption and scattering effects from the surface, and thus, are particularly useful in atmospheric correction and scene generation [55]. For a Lambertian and homogeneous surface with reflectance ρ, a TOA radiance spectrum (L) can be calculated through (1)

\[
L = L_0 + \frac{(E_{dir}\mu + E_{dif})(T_{dir} + T_{dif})\rho}{\pi(1 - S\rho)}
\]

(1)

where \(L_0\) (atmospheric reflected radiance), \(E_{dir/dif}\) (at-surface direct/diffuse solar irradiance), \(S\) (spherical albedo) and \(T_{dir/dif}\) (upwelling direct/diffuse target-to-sensor transmittance) are the atmospheric transfer functions, and \(\mu_{li}\) is the cosine of the solar zenith angle (SZA).

B. Modtran

MODTRAN (http://modtran.spectral.com/) is a widely used atmospheric RTM with multiple applications in Earth observation [13]. It solves the radiative transfer equation with
accurate modeling of the coupled scattering and absorption effects. At its core, MODTRAN implements the discrete ordinates algorithm [56] and the correlated-k method for simulating the gaseous absorptions [57]. The spherically symmetric atmosphere is constituted of stratified vertical profiles of temperature, gas molecules standard, and user-defined aerosol optical properties are distributed in the boundary-layer (<2 km) and stratosphere. Accordingly, MODTRAN propagates the incoming solar irradiance through the Earth’s atmosphere, modeling the effects from spherical refraction, molecular and aerosol absorption/emission and scattering, and surface reflections and emission. The calculated spectral outputs include spectral transmittance, TOA radiance, solar irradiance, and so on. These outputs cover the full optical range (0.2–200 μm) with a resolution up to 0.1 cm\(^{-1}\) for the ultrafine correlated-k simulations, or even higher using the latest line-by-line capabilities [58]. MODTRAN has been extensively validated, and it continues to be maintained and updated.

C. ARTMO’s Emulator Toolbox

As part of the ARTMO software package, the Emulator toolbox enables the evaluation of regression algorithms on their capability to approximate RTM outputs as a function of input variables [24], [37]. Essentially, the emulator toolbox encompasses a suite of methods from the simpleR library [46] that can be combined with DR methods (e.g., PCA) in order to train statistical models that produce spectral outputs based on RTM inputs. While in earlier versions, the focus of the Emulator toolbox was to train emulators based on the vegetation leaf-canopy RTMs available within ARTMO, in the here latest presented version (v. 1.13), one of the introduced novelties involved the ability to handle the multiple RTM spectral outputs stored in the ALG files (i.e., the six atmospheric transfer functions). Accordingly, users can easily develop emulators of atmospheric RTMs and export them for their later use in data processing applications.

IV. EXPERIMENTAL SETUP

A. Database Description

The emulation experiments and assessment are based on a set of atmospheric transfer functions databases that were first generated with ALG using MODTRAN6 and the interrogation technique in [16]. To do so, the input variable space was sampled using an LHS strategy with minimum and maximum boundaries, as given in Table I. These input variables are typically used in the atmospheric correction of optical satellite data due to their influence in shaping the TOA radiance [24].

An LHS of training data is preferred over a systematic grid-ded sampling, as LHS ensures that the set of pseudorandom values cover the full variability of the input parameter space while minimizing the number of samples. Thus, in principle, the developed emulator will be able to reconstruct accurately the spectral atmospheric transfer functions for any possible combination of input variables. The training database size is known to influence the emulator accuracies as it serves to sample the input variable space. Earlier studies suggest that using just 400 samples would suffice to function as an emulator [59], though larger databases may also improve the accuracy of an emulator [37]. Nevertheless, an increasing number of samples also results in longer computation times and can also put a problem due to RAM memory limitations in training the emulators. In the analysis presented here, we built training databases with 100, 500, 1000, and 2000 samples to study the effect of database size in the accuracy while still having an acceptable compromise regarding the computation time and memory constraints.

Regarding the spectral configuration, all these training data sets were generated in the 400–2500-nm spectral range avoiding the saturated H\(_2\)O absorption bands at ∼1380 and ∼1880 nm. The correlated-k option was used for the following three spectral resolution: coarse (15 cm\(^{-1}\)), medium (5 cm\(^{-1}\)), and fine (1 cm\(^{-1}\)). With these resolutions, an entire spectrum consists of nearly 1400, 4000, and 20000 spectral channels, respectively. Simulations at ultrafine spectral resolution (0.1 cm\(^{-1}\)) were discarded due to their high memory requirements (200000 spectral channels, resulting in more than 300-Gb matrices to perform GPR emulators). An increasing spectral resolution increases the complexity of the data in the absorption features (see Fig. 1), which can potentially impact the accuracy of the DR method and, thus, of the emulator.

The processing time of the 1000-samples databases took 3.5 h, 12 h, and 44 h, respectively, for the three increasing spectral resolutions on a personal computer with the following characteristics: Windows 10 64-bits OS, i7-4710 CPU 2.50 GHz, 16-GB RAM, and using five parallel executions.

<table>
<thead>
<tr>
<th>Model variables</th>
<th>Units</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>O(_3) column concentration (O3)</td>
<td>[ant-cm]</td>
<td>0.25</td>
<td>0.45</td>
</tr>
<tr>
<td>Columnar Water Vapour (CWV)</td>
<td>[g·cm(^{-2})]</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Aerosol Optical Thickness (AOT)</td>
<td>unitless</td>
<td>0.05</td>
<td>0.5</td>
</tr>
<tr>
<td>Asymmetry parameter (g)</td>
<td>unitless</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>Ångström exponent (α)</td>
<td>unitless</td>
<td>0.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Single Scattering Albedo (SSA)</td>
<td>unitless</td>
<td>0.7</td>
<td>1</td>
</tr>
<tr>
<td>Surface elevation (h)</td>
<td>[km]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Solar Zenith Angle (SZA)</td>
<td>[deg]</td>
<td>20</td>
<td>70</td>
</tr>
</tbody>
</table>

Fig. 1. Sample of the direct upward transmittance in the O\(_2\)-A absorption band at coarse (yellow), medium (orange), and fine (blue) spectral resolutions.
out of eight cores. The same computer was used to carry out the application scenarios described in Section IV-D.

B. Methodology for Emulator Analysis

The methodology for emulator analysis follows our previous work in [26]. The selected regression algorithms (see Table II and Section II-A) were first analyzed for the default training database (i.e., 1000-samples and medium resolution) for emulating the six spectral atmospheric transfer functions.

The PCA method was first applied for reducing the dimensionality of training data. Although the first five components explain 99.95% of the variance in the medium spectral resolution database, developing the regression models with less than ten components led to inaccurate reconstructions of the atmospheric transfer functions. Higher accuracy can be achieved, particularly in absorption bands, adding more components in the regression algorithms but at the expenses of slowing down processing time. Therefore, the first analysis was carried out with ten components (i.e., 99.99% explained variance) in order to balance accuracy and processing time.

After identifying the best algorithm, the role of training database size was second analyzed. Here, the 100-, 500-, 1000-, and 2000-samples databases at medium spectral resolution were subsequently passed to ARTMO’s Emulator toolbox for the best performing emulator and 10 PCA components.

The role of the DR method and the number of components was third analyzed for the best performing regression algorithm. The PCA and PLS methods were first compared in the default training database in order to identify the best performing method. Second, an increasing number of components (10, 20, 30, and 40) were evaluated for all the six spectral atmospheric transfer functions. The analysis is carried out for the 500-samples databases and for the coarse, medium, and fine MODTRAN6 spectral resolutions.

C. Emulation Validation

Emulators are approximations of the original model, and as such, this approximation introduces a source of error in the emulated spectral outputs [23]. Therefore, validation of the generated emulator is an important step in order to evaluate its accuracy. Two steps were used to validate the accuracy of the emulator. In the first step, a verification was carried out in order to provide a first quick analysis of emulator performance. Through this verification, we determined which regression algorithm methods and training configuration were worth exploring in more detail. For this verification step, the original data were split into two parts following the approach in [24]. Here, a single-random split was applied using 70% samples for training and the remaining 30% for verification. However, it is more suitable to perform alternative cross validation methods (e.g., leave-one-out) to perform the validation when the number of samples in the database is small. In the second step, a performance assessment was carried out in order to compare the accuracy of the various emulators against each other. Here, a reference database of 2000 samples was generated with a random LHS distribution and the same input variables and ranges, as in Table I. This reference database of RTM outputs is used as ground truth to evaluate and to compare the performance of all emulators. Here, we used the spectral normalized root-mean-square-error (NRMSE) [%] (see (2)) to evaluate the performance.

$$\text{NRMSE} = 100 \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f(x_i) - \hat{f}(x_i))^2}$$

where $n = 2000$ is the number of samples in the reference database, $f$ and $\hat{f}$, respectively, the emulated and RTM spectral values evaluated at the input point $x_i$, and $f_{\text{max}}$ and $f_{\text{min}}$ are, respectively, the maximum and minimum values of the $n$ spectra in the reference data set. In order to inspect the emulators’ accuracy and compare their performances along with the spectral range, the spectral NRMSE is plotted as a function of wavelength. The spectrally averaged NRMSE ($\text{NRMSE}_{\text{spa}}$) and the emulation time of the $n$ samples are also tracked.

D. Application Example: Atmospheric Correction

As a proof of concept, the best evaluated emulator is used in the context of atmospheric correction for three scenarios: 1) synthetic hyperspectral TOA radiance spectra as would be observed by an imaging spectrometer overflying a vegetated surface; 2) a real-OLCI/Sentinel-3 data; and 3) a synthetic FLORIS/FLEX data. Before testing the emulators for atmospheric correction, we first evaluated the accuracy of atmospheric RTM emulators in reconstructing TOA radiance spectra. A simulated TOA radiance spectra database ($L_\text{sim}$) was constructed with (1) using the MODTRAN6 simulations in the reference database at medium spectral resolution. Only the conifer spectrum from MODTRAN6’s albedo database was used as Lambertian surface reflectance spectrum in the TOA radiance simulations in order to vary only the input atmospheric parameters. In parallel, an equivalent TOA radiance database ($L_\text{emu}$) was constructed using the atmospheric transfer function emulators. The spectral relative error, in absolute values, between $L_\text{sim}$ and $L_\text{emu}$ was calculated for all the 2000 samples and its histogram plotted against wavelengths. In order to better visualize the errors and avoid the deep H$_2$O absorption regions, the results will be shown in the 400–1400-nm spectral range. For consistency with the emulation validation, the spectral NRMSE will also be provided. We can compare these results against typical absolute radiometric calibration performance for satellite instruments (e.g., 1%–2% in OLCI/Sentinel-3 [60]) and signal-to-noise requirements for imaging spectrometers (i.e., ~0.1%).
Second, we applied the emulator to invert the surface reflectance from the forward TOA radiance simulation. In this case, we started with the aforementioned simulated TOA radiance database, $L_{\text{um}}$ and derived the surface reflectance by inversion of (1) using the emulated atmospheric transfer functions. This produced a database of 2000 inverted surface reflectance spectra, for which we calculated the relative error against the ground-truth conifer reflectance spectrum. The histogram of relative errors is plotted for each wavelengths.

With the aim of testing the utility of emulators in practical applications, as a proof of concept, third, we applied the best emulators to the atmospheric correction of OLCI/Sentinel-3 Level-1B. OLCI is an optical instrument on the board of the Sentinel-3 satellite, belonging to the space segment of the Copernicus program. OLCI acquires the Earth-reflected radiance with 21 spectral channels distributed in the 400–1000-nm spectral range and with a resolution of 2.5–40 nm. Five cameras scan the Earth surface at 300-m spatial resolution with a total swath of 1270 km. More details of the instrument configuration can be found in ESA's Sentinel online https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-olci/olci-instrument/spectral-response-function-datawebsite. We performed the atmospheric correction on a subset OLCI/Sentinel-3B Level-1B TOA radiance data (see Fig. 2).

A subset of nearly 1.4 million pixels was taken within columns 3138–4062 and lines 1500–3000 corresponding to the mountainous area of the Swiss Alps in the center, the coastal area around Nice, France, in the bottom-left, and North–West Italy at the left-hand side of the image. The surface elevation ranges from sea level up to 4 km (1 km on average). The average geometric conditions are 5° of viewing zenith angle and 28° of SZA. As for the atmospheric conditions over the image, the ozone (0.308–0.327 atm-cm) and water vapor (0.47–2.87 g-cm$^{-2}$) are provided by the ECMWF near-real-time data sets and stored in the OLCI Level-1B meteorological data file. The aerosol parameters are extracted from the OLCI Synergy product. For the selected image subset, the average Aerosol Optical Thickness (AOT) and Angstrom exponent values are, respectively, 0.39 and 1.085. For the remaining aerosol parameters (asymmetry factor, $g$, and SSA), we assigned the spectrally averaged values from OPAC's pure water-soluble aerosol type at 50% relative humidity [61] as identified by the synergy algorithm (in the variable called AMID), i.e., $g = 0.67$ and SSA = 0.95. The atmospheric correction was performed by inverting the surface reflectance from (1). The atmospheric transfer functions were first derived pixelwise by running a previously trained emulator with the input geometric and atmospheric conditions of the image. These high spectral resolution spectra were then convolved by the nominal OLCI/Sentinel-3B spectral response function extracted from ESA's Sentinel online https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-olci/olci-instrument/spectral-response-function-datawebsite. Lastly, we analyzed the distribution of surface reflectance values from the emulation inversion and from the OLCI’s synergy surface reflectance product, excluding the O$_2$ absorption (bands #13–#15) and the deep H$_2$O absorption (bands #19 and #20).

As a final application example, optimized emulators were generated for the atmospheric correction of synthetic FLORIS/FLEX data. FLEX is the future's ESA's Earth Explorer mission, aiming at obtaining global maps of vegetation Sun-induced fluorescence emission and biophysical parameters from measurements in the 500–780-nm spectral range [62]. FLEX is equipped with imaging spectrometers (FLORIS) that will measure radiance at a spectral sampling (resolution) up to 0.1 nm (0.3 nm). Since the expected launch date of FLEX is 2023, here, we tested the emulators for atmospheric correction over a synthetic noise-free scene generated with the FLEX end-to-end mission performance simulator [9], [63]. The scene of 100 × 100 pixels (see Fig. 3) consists of 12 homogeneous land cover classes characterized by constant values of leaf area index, chlorophyll content, and background bare-soil surface reflectance (dark/bright), for which the surface reflectance and fluorescence emission are simulated with SCOPE RTM. The homogeneous atmosphere was simulated with MODTRAN6 with a mid-latitude summer profile, a CWV = 2 g-cm$^{-2}$ and aerosol conditions described by an AOT = 0.25, $g = 0.81$, $\alpha = 1.54$, and SSA = 0.98. As for the geometric conditions, the average SZA is 40°, and surface elevation at 0 km.

![Fig. 2. Color composite of OLCI Level-1B data acquired over France on the June 17, 2019, at 10:03. The processed data subset is marked by the red square.](image1)

![Fig. 3. RGB composite (640/550/500 nm) of FLORIS Level-1B TOA radiance simulated with FLEX-E (left) and sample spectra for the twelve distinctive land cover classes (right).](image2)
The emulators for the atmospheric transfer functions were trained with the best performing algorithm, database size, DR method, and a number of components as derived from the analysis described in Section IV-A. The training database consists of the same parameter configuration as in Table I and a spectral configuration optimized for the FLORIS instrument (see Table III). In order to perform the atmospheric correction of FLORIS data, the high spectral resolution emulated atmospheric transfer functions were convolved by a super-Gaussian approximation of FLORIS spectral response [64] following the approach proposed in [65].

V. Results

Results are organized as follows. First, the key aspects that play a role in the emulation of spectral atmospheric transfer functions are systematically analyzed. Then, the best performing emulators are used for the atmospheric correction of synthetic and real satellite data sets.

A. On the Role of Used Regression Algorithms

The performance of the selected regression algorithms for emulating atmospheric transfer function spectra is first analyzed for the medium spectral resolution and using a training database of 1000 samples according to a 70%/30% training/validation distribution and a 10 PCA components reduction. Performance assessment results are compared by plotting the spectral relative errors (NRMSE) (Fig. 4), and these error statistics were averaged for the six spectral variables. In agreement with our previous findings in [24], the results in Fig. 4 indicate that KRR and GPR methods systematically emulate more accurately the six atmospheric transfer functions with average NRMSE values of 0.4% and 0.2%. The other regression algorithms get higher average NRMSE values of 1.9% for NN and 3.5% for RF. Except for the diffuse transmittance ($T_{\text{dir}}$), KRR and GPR show a nearly overlapping spectral NRMSE with an overall value below 1% in most spectral channels. It is also observed that the errors increase inside the deep H2O and O2 absorption bands. This is a result of two combined effects. On the one hand, the DR through PCA is challenged by the reconstruction of all the spectral features in these absorption regions and for the broad simulated spectral range. On the other hand, the nature of the relative error metric implies divisions by nearly zero values in these deep absorption regions, and thus, increases the error values at these wavelengths. When observing the performance for each atmospheric transfer function, the results indicate that the direct components of the upward transmittance ($T_{\text{dir}}$) and the at-surface solar irradiance ($E_{\text{div}}$) are more accurately emulated than their diffuse counterparts ($T_{\text{div}}$ and, specially, $E_{\text{dir}}$) for the GPR and KRR emulators. As for the computation time, all regression methods emulate the 12000 spectra (i.e., the six transfer functions and 2000 conditions of the reference database) in less than 1 s without significant differences. Altogether, GPR seems the optimal method for emulating MODTRAN data followed by KRR. The RF and NN regression methods obtain higher NRMSE values and can, therefore, be discarded to function as an emulator.

B. On the Role of Database Size

The role of database training size was subsequently analyzed. The spectral NRMSE performance evaluation results are shown in Fig. 5 only for the path radiance ($L_0$) and the at-surface diffuse solar irradiance ($E_{\text{dir}}$) outputs for illustration purposes. Similar results are obtained for the remaining atmospheric transfer functions (figures not shown). These results indicate how increasing the database size from 100 to 2000 samples (70% used for training) reduced the relative errors to nearly half along with the entire spectral range. However, the error reduction is not systematic with the addition of more samples in the training database. Indeed, the errors are largely reduced when passing from 100 to 500 samples but they get more stabilized for larger training databases (i.e., 500, 1000, and 2000 samples). In fact, the performance of the $L_0$ emulator seems to give slightly better results with 500 training samples than with 2000 training samples.

Table IV summarizes the spectrally averaged NRMSE values as a function of the database size for all the atmospheric transfer functions. It is observed again how the error is reduced from 0.62% to 0.18% when passing from 100 training samples to 2000 samples. The “saturation” effect in the performance for 500–1000 training samples is also observed for all the atmospheric transfer functions. Regarding the emulation computation time, the average CPU time spent to emulate the $n = 2000$ spectra from the reference database shows a nearly proportional dependence of the training database size.

<table>
<thead>
<tr>
<th>Function</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0$</td>
<td>0.83</td>
<td>0.30</td>
<td>0.30</td>
<td>0.28</td>
</tr>
<tr>
<td>$E_{\text{dir}}$</td>
<td>0.35</td>
<td>0.14</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>$E_{\text{div}}$</td>
<td>0.58</td>
<td>0.32</td>
<td>0.31</td>
<td>0.33</td>
</tr>
<tr>
<td>$S$</td>
<td>0.45</td>
<td>0.16</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>$T_{\text{dir}}$</td>
<td>0.31</td>
<td>0.16</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>$T_{\text{div}}$</td>
<td>1.16</td>
<td>0.36</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Mean:</strong></td>
<td>0.62</td>
<td>0.24</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>CPU (s):</strong></td>
<td>0.14</td>
<td>0.27</td>
<td>0.50</td>
<td>1.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda$ (nm)</th>
<th>FLORIS (MODTRAN) sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>497.500 – 685.750</td>
<td>2.0 nm (1.0 cm$^{-1}$)</td>
</tr>
<tr>
<td>685.750 – 697.125</td>
<td>0.1 nm (0.1 cm$^{-1}$)</td>
</tr>
<tr>
<td>697.125 – 739.375</td>
<td>1.0 nm (1.0 cm$^{-1}$)</td>
</tr>
<tr>
<td>739.375 – 780.625</td>
<td>0.1 – 0.5 nm (0.1 cm$^{-1}$)</td>
</tr>
</tbody>
</table>
C. On the Role of Dimensionality Reduction Methods

The influence of DR methods on the emulator accuracy is assessed on the best performing regression algorithm, i.e., GPR. For a GPR emulator and ten components of DR, the performance of PCA and PLS transformation is shown in Fig. 6. Here, only the emulator for the at-surface diffuse solar irradiance is shown since the performance results with the PCA and PLS methods show larger differences. For this specific atmospheric transfer function, the PCA method is the best performing. However, for the remaining functions, the difference in performance with PCA and PLS methods is less evident as observed in the spectrally averaged NRMSE values in Table V. Among the six atmospheric transfer functions, it is worth noticing that the emulator for $E_{\text{dif}}$ systematically gets the higher errors with both DR methods.

Following these results, the PCA method was chosen to evaluate the role of a number of components used for training the GPR models as a function of the spectral resolution (coarse, medium, and fine). The assessment was carried out by using 10, 20, 30, and 40 components. In Fig. 7, only the results for the upward direct transmittance ($T_{\text{dir}}$) are shown for illustration purposes. Similar results were obtained for the remaining atmospheric transfer functions. The spectral range was selected to illustrate the effect of adding more PCA components in a region affected by several gas absorption features like the H$_2$O (720, 820, and 940 nm) and the O$_2$-A.
A clear trend is observed when comparing the results on the various spectral resolutions; as the resolution increases from coarse to fine, the bias in the NRMSE error (i.e., outside of absorption regions) decreases from 0.2% to 0.05%. However, the error difference between this bias and the values in the absorption regions increases from ~0.1% in the coarse resolution up to ~0.2% in the fine resolution. For Fig. 7(a)–(c), it is also observed that adding more components into the model reduces the relative errors in the absorption regions. Nevertheless, there is a saturation effect after 20–30 PCA components, where the NRMSE error does not decreases neither in its absolute nor random values. The drawback of adding more components is increasing the model complexity, and thus, a longer computation time. For instance, emulating 2000 spectra in the fine spectral resolution increases from 1.8 s using ten components to 3.6 s using 40 components.

D. Application of Emulators in Atmospheric Correction

As a first demonstration case, we compared the TOA radiance spectra constructed from MODTRAN simulations against those constructed from the best performing emulations. To do so, the medium resolution GPR emulators trained with 500 samples and 20 PCA components were run at the input conditions in the reference database. The histogram of relative errors between simulated and emulated TOA radiance is given in Fig. 8 in the log-scale together with the average relative error.
Fig. 8 shows that, outside of main absorption bands, the TOA radiance is reconstructed with an error <0.6% for 95% of the conditions in the reference database. These errors are even lower (~0.15%) in average atmospheric/geometric conditions and even down to ~0.01% for the best 10% of the 2000 simulations. When analyzing the errors inside of gas absorption regions, they increase due to the divisions by nearly zero radiances. The upper 95% percentile of the errors increases above 1%, with values of 2% in the O_2-A at 761 nm and up to 10% in the deep H_2O at 1125 nm. However, the average values are typically below 1%. For consistency and comparison with the performance metrics shown along the previous sections, the spectral NRMSE is shown in Fig. 9.

The same analysis was carried out in terms of the inverted surface reflectance. The histogram of the relative errors (Fig. 10) shows an increase toward the 400–500-nm spectral range due to the low surface reflectance values and the higher impact of the aerosol scattering. Nevertheless, the errors are rather spectrally flat after 600 nm (excluding the absorption regions). The relative error is below 1%, when excluding deep absorption regions, for 95% of the atmospheric conditions and up to 3% in the O_2-A and the H_2O band at ~820 nm.

In our second test case scenario, the 500-sample GPR emulators of MODTRAN atmospheric transfer functions at medium spectral resolution were applied to atmospherically correct a subset of OLCI Level-1B data. The processing time was 25 m for 1.3 million pixels in batches of 100,000 pixels to avoid saturating the RAM memory. Fig. 11 shows the statistics of spectral reflectance values in the OLCI image as provided in the OLCI synergy product (blue) and as obtained with our emulators (red). Fig. 11 indicates a generally good agreement on the derived image values. However, some discrepancies are visible in the two first spectral channels, where the choice of the extraterrestrial solar irradiance spectrum and aerosol optical properties have a higher impact. Above 800 nm, the dynamic range of the values obtained by the emulator shows a tendency to obtain lower reflectances than those in the OLCI synergy product, with differences around 5%.

In Fig. 12, we show a sample of five surface reflectance spectra covering the dynamic range observed in the image, i.e., from low to high values. Fig. 12 shows a good agreement between the two atmospherically corrected reflectance spectra.

In our last test case scenario, synthetic FLORIS/FLEX data were atmospherically corrected using MODTRAN-based emulators. Querying the emulator for the input conditions of the 10,000 image pixels took less than 30 s. Sample surface reflectance spectra for the 12 distinctive land cover classes in Fig. 3 are shown at the bottom of Fig. 13. An overall agreement is found for all distinctive surfaces, with errors ~1% inside the O_2-A and 3.6% in O_2-B absorptions. These errors are slightly above the FLEX mission requirements (1%). However, it must be remarked that the error propagation combines the emulator and the residual spectral calibration errors.
the O$_2$-A (left-hand side) and O$_2$-B (right-hand side) absorption bands.

Fig. 13. Sample surface reflectance spectra from the ground truth reference (blue) and retrieved with the emulators (red). Bottom: zoom window in the O$_2$-A (left-hand side) and O$_2$-B (right-hand side) absorption bands.

VI. DISCUSSION

In this section, we discuss the most important messages derived from analyzing the emulator results (Section VI-A), and the implications on further studies for atmospheric correction (Section VI-B). We finally discuss possible data processing opportunities with atmospheric emulators (Section VI-C).

A. Interpreting Emulator Results

Emulation is a statistical technique used to approximate deterministic models of large computational burden [23]. Previous experiments have shown that regression algorithms can accurately function as emulators to approximate RTMs [20], [24], [37] and obtain competitive results with respect to traditional LUT interpolation [21]. In this work, we explored further the concept of emulating spectral transfer functions for atmospheric correction. MODTRAN6 was used to generate the databases of atmospheric transfer functions, but the emulation technique can be applied to any other atmospheric RTM as demonstrated in [25]. Specifically, key aspects that impact the performance of an emulator were systematically analyzed. These aspects include: 1) used statistical regression algorithm; 2) training database size; and 3) used DR method and number of components, linked with spectral resolution. Each of these aspects is briefly discussed in the following.

The first aspect studied in this work is a comparison of the performance of various regression algorithms. From the five algorithms tested, GPR and KRR reached systematically the best accuracies for all the six atmospheric transfer functions in agreement with our previous results in [24]. The NNs and RFs tested here were not able to reach the accuracy achieved by GPR. Though a multi-output version of the GPR method exists [66], the version used in this work is a single-output method, and thus, looping over each individual PCA component was the employed methodology to recover the multi-output spectral data. Consequently, the training and application of these models will take longer times when more components are added to the emulator. However, GPR has an additional advantage over other methods. While all emulators- and LUT-based methods allow computing derivatives numerically, only kernel methods (KRR and GPR) yield the explicit formulas for the calculation of Jacobians analytically. This allows error quantification and propagation directly from their model structure, avoiding the implementation of complex error propagation strategies nor requiring the calculation of any numerical approximation or empirical derivatives [35], [67]. These associated uncertainty estimates can function to propagate errors from atmospheric correction algorithms and the Jacobians in efficient numerical inversion algorithms. Finally, although the best performing emulators reached accuracies with a relative error (NRMSE) of 0.2% for the tested data sets, there is still a margin for improvement foreseen. The latest generation of deep GPRs is indeed a promising alternative to the classical GPRs used in this work [68], [69]. Moreover, we can think about constraining the emulators-based physical correlations of the six atmospheric transfer functions, which were here emulated independently. For example, it is known that an increase in AOT will increase the absorption and the multiple scattering, thus reducing the direct transmittance ($T_{\text{dir}}$) while increasing the diffuse transmittance ($T_{\text{dif}}$). By including these correlations, we would expect better conditioned emulators with an improvement in accuracy and performance.

A second aspect studied in this work is the effect of the training database. Based on the usage of traditional LUT interpolation methods, it was initially expected that larger training databases would turn into more-accurate emulators. This aspect was investigated in order to optimize the emulator accuracy. Yet, the results indicate a residual increase in the GPR emulator accuracy between 500 and 1000 samples and a kind of saturation effect after 1000 samples. One explanation is that adding more samples in the same input variable space does not add new information to the statistical model. With a few samples, the GPR emulator can already recover the smooth dependencies of the atmospheric transfer functions with the input variables. Another explanation is that the reference data set was generated with the same data distribution (LHS) than the training data sets. Thus, the emulator is somehow optimized for variations of the spectral outputs in this input variable space and distribution. Moreover, it must be remarked that larger training databases go along with longer processing times, both in database generation and emulator execution, nearly doubling the computation time when the training database doubled its size. It worth noting that in this work, we did not study the impact of sampling methods in the training database. And only the LHS method was used for its simplicity and accuracy [24]. However, the methods known as experimental design (also known as active learning) can potentially be used to further improve the accuracy of an emulator [70]–[72]. Indeed, these methods aim at finding an optimal (and minimum) set of training samples with a consequent reduction in the error of the statistical regression and its computation time.
A third aspect was the evaluation of DR in terms of algorithms and number of components, and how these DR methods perform with an increasing spectral resolution. The implementation of the DR step is key for the emulation of multi-outputs models, as it allows reducing the RTM spectral output to a manageable number of components before applying a single-output regression algorithm. A suitable DR method should also allow reconstructing the full spectrum in order to deliver spectroscopy outputs of several thousands of bands, and thus, limit the choice of other more advanced DR methods [73], [74]. In this work, we analyzed two classical statistical approaches being PCA and PLS. On the one hand, the PCA method solves the multicollinearity problem in the model outputs without considering the correlations between dependent (outputs) and independent (inputs) variables. On the other hand, PLS is a more advanced version of PCA that takes into account these correlations between variables. However, in the absence of clusters in the input variables, both methods are mathematically alike which explains why PLS did not lead to performance improvement as compared with PCA for the analyzed number of components. In fact, the PCA method derived slightly lower errors in the reconstruction of the six atmospheric transfer functions and for all spectral channels. Once selected the PCA method to train our GPR emulators, we analyzed the effect of increasing the number of components. Indeed, we could expect that by adding more components in the DR we would achieve better spectral reconstruction, particularly at higher spectral resolution. Previous research in [48] and [49] indicate that 100–400 PCA components are needed to accurately reconstruct the spectral data. Yet, our results indicate that adding more than components 20–30 did not seem to further reduce the reconstruction errors. The same conclusion is derived for the coarse, medium, and fine spectral resolutions, though with small error differences when comparing inside and outside absorption regions. The apparent discrepancy can be explained by the application domain and error requirements. The cited works focus on reproducing high spectral resolution (0.5 cm\(^{-1}\)) spectra from the IASI/MetOp instrument inside of deep gas absorptions for applications in trace gas monitoring, where a selection of the number of components was based on having a reconstruction error below the random instrument noises. We instead focus on a broader spectral range and coarser spectral resolution for applications in atmospheric correction, choosing a more relaxed error requirement based on absolute radiometric calibration. This comparison between works underlines that the shape of the spectral data, application domain, and error requirements determine the optimized number of components. In practice, some repetitions are required to deduce this optimal number. Nevertheless, although PCA can be concluded as the preferred DR method among the two explored options, the development of alternative DR methods to improve the reconstruction of the signal in the entire spectral range is strongly encouraged (e.g., [50]). An envisaged option is to consider the known physics of the radiative transfer in a DR method. Here, we could decouple the gas absorption from the molecular and aerosol scattering following a similar approach to [75]. For each of the six atmospheric transfer functions \(T_s\), we would have an effective gas absorption \(T \times T_{gas,x}\) whose spectral features are well known and common for all of these spectral functions based on the formulation of the spectral optical depth \(\tau_i\) in [76]. An effective optical depth (per gas molecule) could be used as a DR component to account for variations in the gas concentration, pressure conditions, and optical path due to coupled absorption-scattering effects. In addition, the molecular (Rayleigh) and aerosol scattering consist of smooth spectra that can be empirically represented by a double exponential or by a simplified physical model, such as [12]. These empirical parameters correspond to new DR components. This concept is represented in (3)

\[
T_s \equiv T_{scat}(p_s) \cdot T_{gas,x} = T_{scat}(p_s) \cdot \exp \left( \sum_{i=1}^{n_g} w_{i,x} \tau_i \right) (3)
\]

where \(p_s\) are the parameters of the scattering curve and \(w_{i,x}\) are the parameters modifying the effective optical depth of the \(n_g\) gases, both specific for each transfer function \((x\ subindex)\). The gaseous spectral optical depth is provided by running an atmospheric RTM with fixed atmospheric conditions.

B. Opportunities for Atmospheric Correction

The applicability of emulating atmospheric RTM data has been demonstrated in three study cases of atmospheric correction of real and synthetic satellite data. The first test case served to validate the correct functionality of the trained GPR emulators over a synthetic data set of MODTRAN TOA radiance simulations. The reconstruction of simulated TOA radiance spectra was achieved with an average (maximum) error of \(\sim 0.1\% (0.5\%)\), excluding deep H\(_2\)O absorption bands, in a wide range of atmospheric and geometric conditions. These errors are an order of magnitude lower than current state-of-the-art absolute radiometric calibration performances of satellite spectrometers [60], and on the same order of signal-to-noise requirements for imaging spectrometers. The emulator performance was further analyzed in terms of atmospherically corrected surface reflectance data. Here, the results showed errors typically below 1%. These errors are in agreement with our previous analysis in [21], demonstrating the potential accuracy of emulators against classical LUT interpolation for atmospheric correction.

In the second test case, the atmospheric correction of nearly 1.3 million pixels of OLCI L1B data subset took about 25 m, showing that GPR emulators can be used efficiently in practical applications. In terms of computation time and memory usage of the GPR emulators, further improvements are expected when optimizing the number of spectral channels and resolution of the simulated training database. Moreover, if the emulators are trained with the simulated data after spectral convolution by the ISRF, the emulation time will drastically drop down when reducing several thousand spectral channels from the high spectral resolution data down to 21 spectral channels of OLCI data. In addition, this pre-processing step of spectral convolution will avoid the need of applying a DR technique, consequently saving computation time and improving the emulator accuracy. However, this strategy might still increase the errors in case of a strong
smile effect in an imaging spectrometer, particularly within absorption bands. In terms of accuracy, the discrepancies in the retrieved surface reflectance statistics can be explained due to several reasons. First, the OLCI SYN surface reflectance product is obtained using the MOMO RTM [14] to perform the atmospheric correction [6], while we used MODTRAN to train the GPR emulators. The differences in the radiative transfer solver (e.g., vectorial in MOMO and scalar in MODTRAN) and in the aerosol parameters (OPAC aerosols in [6] against effective optical properties in our MODTRAN GPR emulators) might explain the statistical differences, particularly at shorter wavelengths where the aerosol scattering and atmospheric polarization of the signal have a stronger effect. Second, the atmospheric correction in the OLCI SYN processing is carried out at TOA reflectance data, i.e., after normalization by the extraterrestrial solar irradiance measured by the solar diffuser on-board of Sentinel-3. We instead processed the OLCI data in radiance units, using the default extraterrestrial solar irradiance spectrum in the MODTRAN simulations. The choice of the solar irradiance spectrum could explain differences along with OLCI’s spectral range. Third, the first OLCI band (centered at 400 nm with a spectral resolution of 15 nm) covers observations below the spectral range in our training database (400–2500 nm). At these short wavelengths, the deep O$_3$ absorption and spectral shape of the solar irradiance spectrum are not well covered in our simulations, which explains the negative reflectances obtained for the first OLCI spectral channel. Finally, our processing with emulators assumes ideal Lambertian surface reflectance and neglects the adjacency and topographic effects, which are likely important in the selected mountainous area.

In the final test case, we trained GPR emulators from an MODTRAN database of atmospheric transfer functions tailored for the hyperspectral FLORIS/FLEX data. The FLEX mission simulator was used to generate realistic instrument FLORIS data over various vegetated and bare-soil surfaces and standard atmospheric conditions. In terms of computation efficiency, the atmospheric correction with GPR emulators took less than 30 s, a fraction of the several minutes that might have taken with a classical LUT interpolation technique [21]. In terms of accuracy, the errors in surface reflectance were $\sim$1% (4%) in the O$_2$-A (O$_2$-B) absorption band. These errors are slightly above the mission requirements for accurate retrieval of Sun-induced fluorescence emission [62]. We have to remark here the stringent reflectance error requirements in a challenging spectral region affected by narrow, spectrally variable, and deep O$_2$ absorption features. Indeed, along with this work, we have observed that the trained emulators typically achieve lower accuracies inside of gaseous absorption regions than in spectrally smooth regions. This indicates that developing more accurate DR methods (statistical-based or physically based) should be an important research line to improve the accuracy of emulators for atmospheric correction.

C. Data Processing Applications With Atmospheric Emulators

Emulators offer similar and additional advantages when compared with traditional LUT interpolation techniques. First, there is a tremendous gain in processing speed when comparing emulators to the original RTM simulations. Second, emulators have competitive or superior accuracies to classical LUT interpolation techniques. Third, as with any regression method, GPR emulators provide associated uncertainties that are not straightforward to calculate with LUT interpolation methods. Fourth, Jacobians are directly calculated based on the analysis of the numerical derivatives in the LUT nodes. Finally, the memory consumption is much lower than storing and reading a large LUT for its interpolation. Consequently, RTM emulators can potentially be applied in a diversity of remote sensing (RS) applications. Some of these applications were already described in [26], such as fast computation of global sensitivity analysis [24], [25]. Here, we outline two more promising emulation applications related to atmospheric RTMs.

1) Numerical Inversion: Iterative optimization is a technique to invert geophysical parameters from spectral data by comparing it against RTM simulations. Commonly used in aerosol characterization [6], the optimization consists in minimizing a cost function, i.e., the difference between measured and estimated TOA radiance by successive iterations in the input variables. This iterative procedure can be time-consuming when large spectral data sets are inverted and impractical for computationally intensive RTMs. Therefore, interpolating precomputed LUTs is common practice in atmospheric characterization algorithms [5], [7]. Following the same approach, emulators can be used to replace the original RTMs and be applied in numerical inversion as an attractive alternative to classical LUT-based methods. It would bypass the need to generate large LUTs and develop complex LUT interpolation routines while leading to faster and more accurate results [21].

2) Data Assimilation and Numerical Weather Prediction: Numerical weather prediction relies on the assimilation of satellite TOA radiance measurements [77]. This assimilation is carried out by simulating the processes of emission and absorption by surface, clouds, and gases, along the line-of-sight of satellite instrument measurements. In order not to delay the production of weather forecasting, the RTM simulations must be made in a few milliseconds. Hence, fast atmospheric models are a major requirement to enable the assimilation of satellite measurements within weather prediction models. An example of such RTMs is the RTTOV model [11], a fast model for simulating radiances acquired by passive optical and microwave satellite instruments (e.g., IASI). With an increasing spectral resolution of satellite instruments, efficiency and accuracy are current challenges of RTTOV [49]. Atmospheric emulators offer an alternative, or complement, to fast RTMs for data assimilation applications. Moreover, the ingestion of uncertainties on data assimilation schemes is essential; here, the use of emulators could represent an additional advantage.

VII. Conclusion

Emulation is a technique to approximate deterministic models based on statistical regression algorithms. With a competitive or even higher accuracy than traditional LUT interpo-
lation, emulators also provide important savings in memory and in computation time. Hence, they offer practical solutions for data processing applications and new research opportunities by converting computationally expensive RTMs into fast surrogate models. Given a subset of typical input variables used in atmospheric correction, we have first analyzed key aspects of emulator design that play a role in the accuracy of reproducing MODTRAN simulations of spectral atmospheric transfer functions, being: 1) the used statistical regression algorithm; 2) the training database size; 3) the used DR method to simplify the spectral dimension into components; and 4) the number of components to where the regression is applied as a function of the spectral resolution of the simulated data. The GPR was found best suited to function as an emulator for the six atmospheric transfer functions. With a training database of approximately 1000 samples, the GPR emulators can reconstruct MODTRAN spectral outputs for any combination of the input variables with an average NRMSE of 0.2%. From the two tested standard DR methods, PCA achieves slightly better accuracies when decomposing and reconstructing the spectra. Typically, 20 components achieve the lowest errors for various spectral resolutions. A larger number of components might better preserve the spectral variability within gas absorption regions but at the expenses of increasing the computation time. The best performing emulators were applied second in three atmospheric correction scenarios for real-OLCI Level-1B data and synthetic hyperspectral and FLORIS/FLEX Level-1B data. Nearly 1.3 million OLCI pixels were atmospherically corrected in less than 25 m and in less than 30 s for 10,000 FLORIS pixels. In terms of accuracy, the atmospheric correction using emulators was achieved with errors below 1% outside of absorption regions and below 3.6% in the O2 bands for FLEX measurements. These results show that emulators can replace computationally expensive atmospheric RTMs and potentially be applied in atmospheric correction of satellite instruments at various spectral resolutions with a competitive accuracy and computational efficiency. Moreover, emulators can be implemented in any other data processing application, where atmospheric RTMs are involved. Finally, we have identified potential research lines to improve the accuracy of atmospheric RTMs involving the latest progress in reproducing MODTRAN simulations of spectral atmospheric transfer functions, being: 1) the used statistical regression algorithm; 2) the training database size; 3) the used DR method to simplify the spectral dimension into components; and 4) the number of components to where the regression is applied as a function of the spectral resolution of the simulated data. The GPR was found best suited to function as an emulator for the six atmospheric transfer functions. With a training database of approximately 1000 samples, the GPR emulators can reconstruct MODTRAN spectral outputs for any combination of the input variables with an average NRMSE of 0.2%. From the two tested standard DR methods, PCA achieves slightly better accuracies when decomposing and reconstructing the spectra. Typically, 20 components achieve the lowest errors for various spectral resolutions. A larger number of components might better preserve the spectral variability within gas absorption regions but at the expenses of increasing the computation time. The best performing emulators were applied second in three atmospheric correction scenarios for real-OLCI Level-1B data and synthetic hyperspectral and FLORIS/FLEX Level-1B data. Nearly 1.3 million OLCI pixels were atmospherically corrected in less than 25 m and in less than 30 s for 10,000 FLORIS pixels. In terms of accuracy, the atmospheric correction using emulators was achieved with errors below 1% outside of absorption regions and below 3.6% in the O2 bands for FLEX measurements. These results show that emulators can replace computationally expensive atmospheric RTMs and potentially be applied in atmospheric correction of satellite instruments at various spectral resolutions with a competitive accuracy and computational efficiency. Moreover, emulators can be implemented in any other data processing application, where atmospheric RTMs are involved. Finally, we have identified potential research lines to improve the accuracy of atmospheric RTMs involving the latest progress in deep GPRs, implementation of active learning methods, and optimization of DR techniques-based physical approximations. It is expected that with a combination of these improvements, emulators will further increase their accuracy in reproducing the spectral outputs of an atmospheric RTM.

REFERENCES


Jordi Munóz-Marí was born in Valencia, Spain, in 1970. He received the B.Sc. degree in physics, the B.Sc. degree in electronics engineering, and the Ph.D. degree in electronics engineering from the Universitat de Valencia (UV), Valencia, in 1993, 1996, and 2003, respectively.

He is an Associate Professor with the Department of Electronics Engineering, UV, where he teaches machine learning, big data, digital signal processing, and electronics, and also a Research Member of the Image Processing Laboratory. His research interests are tied to machine learning, statistical methods, and digital signal processing applied to data analysis in general and remote sensing (RS) in particular. He is a skilled programmer in different computer languages, such as C/C++, Java, Python, and others. He has authored and coauthored many journal articles, book chapters, and conference papers. Please visit https://www.uv.es/jordi/ for more information.

Neus Sabater received the B.Sc. degree in physics, the M.Sc. degree in remote sensing (RS), and the Ph.D. degree in RS from the Universitat de Valencia, Valencia, Spain, in 2010, 2012, and 2018, respectively.

Since 2012, she has been involved in the activities of the Laboratory for Earth Observation, Image Processing Laboratory, University of Valencia, as a Research Technician. Her main activities during this period were related to the development of the preparatory activities of the FLEX mission. In 2013, she was supported by the Ph.D. scholarship from the Spanish Ministry of Economy and Competitiveness and associated with the Ingenio/Seosat Spanish space mission. Her main research and personal interests include atmospheric correction, atmospheric radiative transfer, meteorology, and hyperspectral RS.

Dr. Sabater received the award of the University of Valencia for the best student records in the M.Sc. degree of RS from 2012 to 2013.
Gustau Camps-Valls (Fellow, IEEE) received the Ph.D. degree in physics from the Universitat de Valencia, Valencia, Spain, in 2002. He is a Full Professor of electrical engineering and a Coordinator of the Image and Signal Processing Group, Image Processing Laboratory, Universitat de Valencia. He is also involved in the development of machine learning algorithms for geoscience and remote sensing (RS) data analysis. He has authored 200 journal articles, more than 200 conference papers, and 20 international book chapters.

Dr. Camps-Valls was a recipient of the prestigious European Research Council Consolidator Grant on Statistical Learning for Earth Observation Data Analysis in 2015. He has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE GEOSCIENCE AND RS LETTERS, and the IEEE SIGNAL PROCESSING LETTERS. He was the Invited Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING in 2012 and the IEEE Geoscience and Remote Sensing Magazine in 2015. He holds Hirsch’s index, $h = 60$ (source: Google Scholar), and entered the ISI list of Highly Cited Researchers in 2011. One of his papers was identified on kernel-based analysis of hyperspectral images as a Fast Moving Front research by Thomson Reuters Science Watch.

José Moreno (Senior Member, IEEE) is a Professor of earth physics with the Department of Earth Physics and Thermodynamics, Faculty of Physics, University of Valencia, Valencia, Spain, teaching and working on different projects related to remote sensing (RS) and space research as responsible for the Laboratory for Earth Observation. He is also the Director of the Laboratory for Earth Observation, Image Processing Laboratory/Scientific Park, Valencia. He has been involved in many international projects and research networks, including the preparatory activities exploitation programs of several satellite missions, such as ENVISAT, CHRIS/PROBA, GMES/Sentinel, and SEOSAT, and the Fluorescence Explorer, European Space Agency (ESA)’s Eighth Earth Explorer mission. His main work is related to the modeling and monitoring of land surface processes by using RS techniques.

Dr. Moreno was a member of the ESA Earth Sciences Advisory Committee from 1998 to 2002. He has been a member of the Space Station Users Panel and other international advisory committees. He has served as Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 1994 to 2000.