IMAGE & SIGNAL PROCESSING (ISP) - UNIV. OF VALENCIA

PHYSICS-AWARE EMULATORS FOR ATMOSPHERIC CORRECTION

J. VICENT SERVERA, L. MARTINO, G. VARANDO, G. CAMPS-VALLS

EMULATION FOR ATMOSPHERIC RTM

INTRODUCTION

Atmospheric radiative transfer models

Atmospheric radiative transfer models (RTMs) simulate the physical interaction of light with the atmospheric constituents. Complex, accurate but slow for operational data processing

Current challenges

- Slow RTM → look-up table (LUT) interpolation
- Large LUTs needed for accurate atmospheric correction of hyperspectral data
- Large runtime & RAM usage + tailored approaches

Emulation approach

Regression through statistical models built from small training datasets. Medium speed, high accuracy. However:

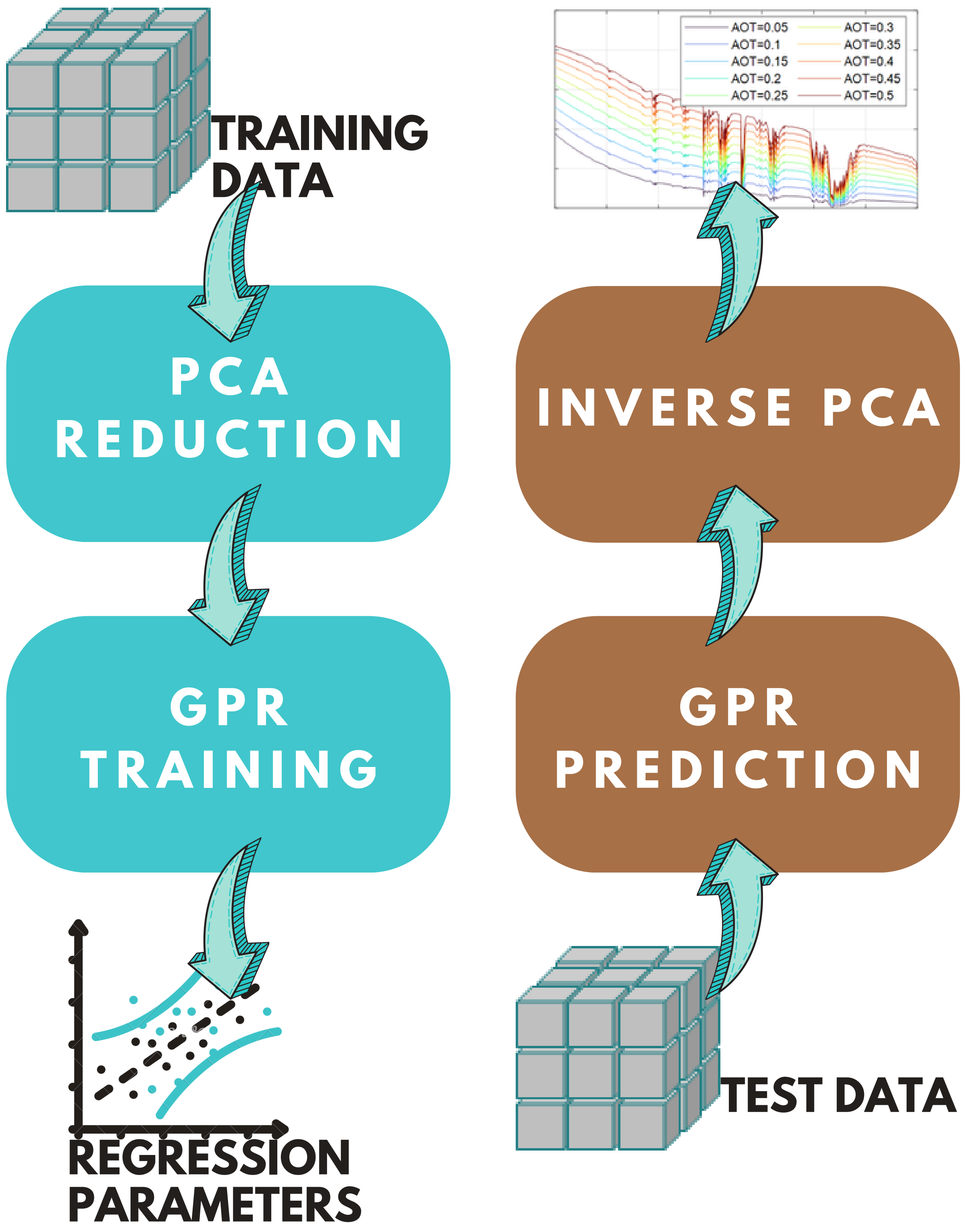
- "Black box" (purely statistical, no physics included)
- Still a challenge for multi-output (spectral) data
- Need for optimization (x5)

THE GP EMULATION APPROACH

CURRENT IMPLEMENTATION

- Each atmospheric transfer function is emulated independently
- Multi-output (spectral) data is handled through PCA dimensionality reduction
- Gaussian Process Regression (GPR) applied on each PCA component. Inverse transformation with the predicted PCA components to recover spectral data
- GPR model (hyperparameters) and PCA vectors stored for later predictions

Question: how to design explainable physics-aware emulators?



FEAT. SELECTION

An iterative process adds new input features in a GP emulator, creating a ranking based on their impact on on model performance.

Algorithm 1 Forward selection

Input: The dataset $\{\mathbf{x}_i, y_i\}_{i=1}^n$, an empty vector $\mathbf{z}_i^{(0)}$, the set of indices $\mathcal{J}_0 = \{1, 2, \ldots, d\}$, and $V(0) = \chi_0 \equiv \left[\frac{1}{n}\sum_{i=1}^n |y_i - \bar{y}|^p\right]^{1/p}$ where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$

for k=1 **to** d **do**

  for $j \in \mathcal{J}_{k-1}$ **do**

    1. Set $\widetilde{\mathbf{z}}_{ij}^{(k)} = [\mathbf{z}_i^{(k-1)}, x_{ij}]$

    2. Compute $\chi_{kj} \equiv \left[\frac{1}{n}\sum_{i=1}^n \left|y_i - g_k(\widetilde{\mathbf{z}}_{ij}^{(k)})\right|^p\right]^{1/p}$

  end for

  3. Set $j^* = \arg\min_j \chi_{kj}$ and $V(k) = \min_j \chi_{kj}$.

  4. Set $\mathbf{z}_i^{(k)} = [\mathbf{z}_i^{(k-1)}, x_{ij^*}]$.

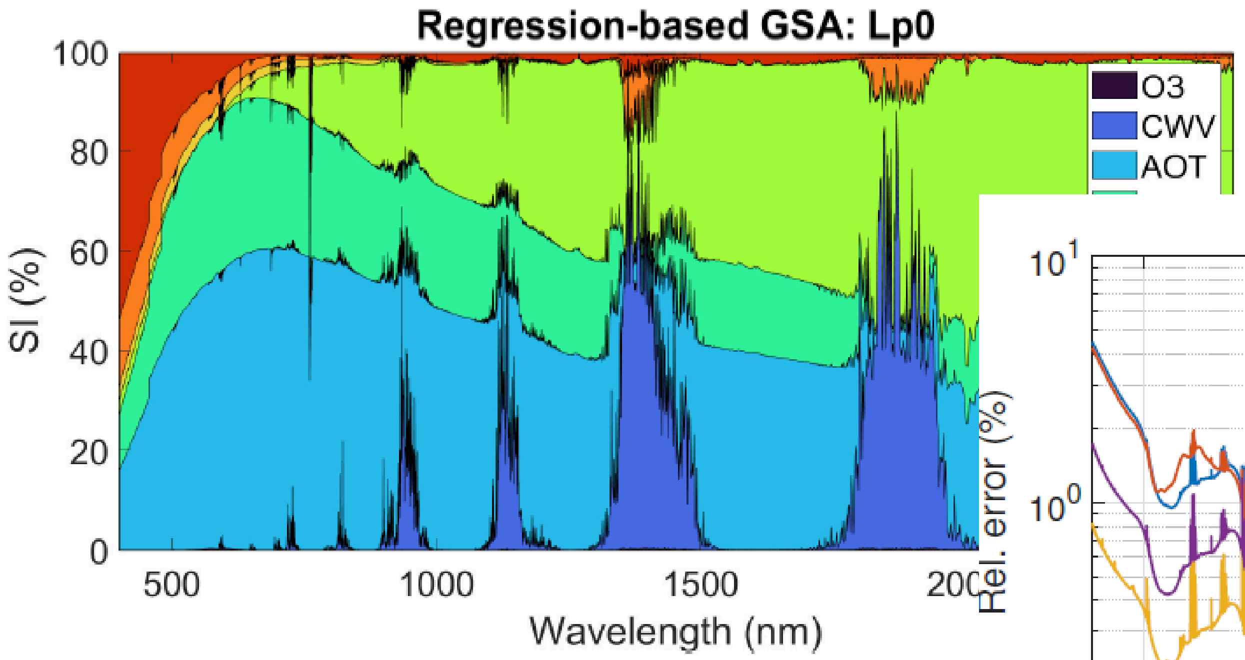
  5. Remove j^* from $\mathcal{J}_{k-1}$ and set $\mathcal{J}_k = \mathcal{J}_{k-1} \backslash \{j^*\}$.

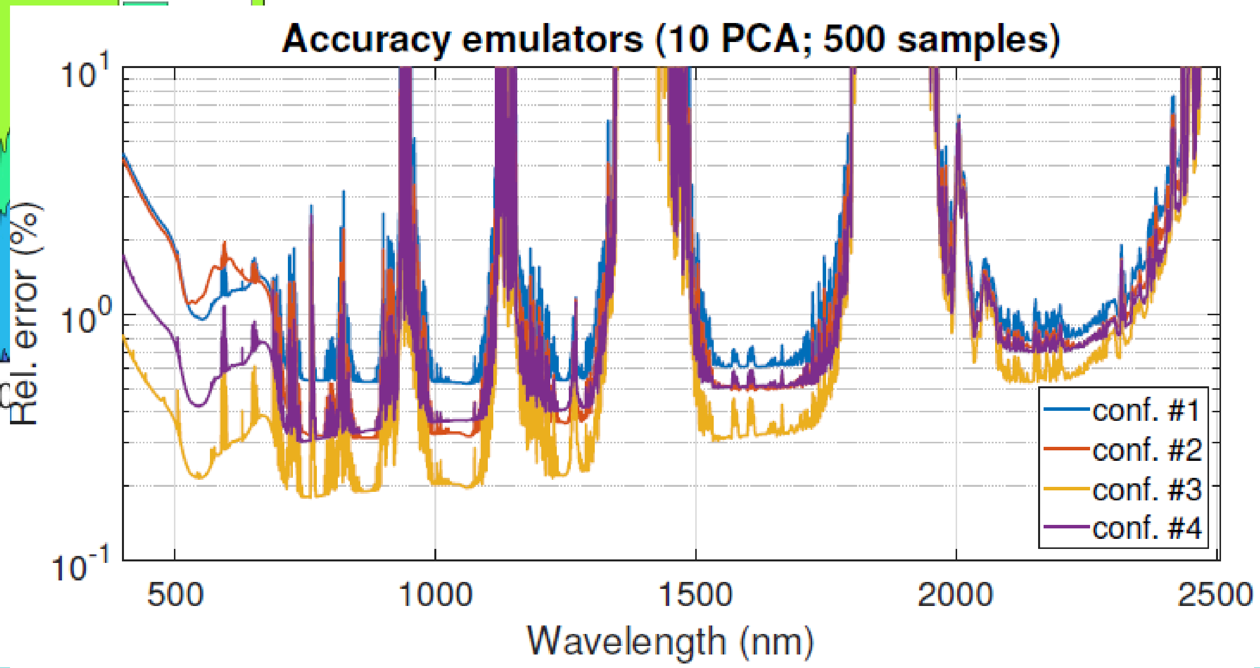
end for

Output: $\mathbf{z}_i^{(d)}$ and $V(k)$ for $k = 0, \ldots, d$

Results and conclusions

- Reducing the number of relevant features makes the emulator physics-aware, improving its accuracy
- The relevant importance of each feature enhances the explainability of the GP hyperparameters
- The feature selection method applied on each wavelength provides a sensitivty analysis of the RTM





SYMBOLIC REGRESSION

Replace PCA by a semi-empirical linear equation. The coefficients of the linear transformation serve as a physics-based dimensionality reduction. LASSO is applied to find the relevant components of the model.

Algorithm 1 Physics-based GP emulator (*training*)

Input: Training dataset $\{\mathbf{x}_i, \mathbf{g}(\mathbf{x}_i)\}_{i=1}^{n}$, basis features (Φ)

$\mathcal{J}_r \leftarrow$ Apply LASSO on a training data subset

Coefficients $\{\alpha(\mathbf{x_i})\}_{\mathbf{i=1}}^{\mathbf{n}} \leftarrow$ Least-squares regression on $\Phi \rightarrow \mathbf{g}(\mathbf{x}_i)$

for c=1 **to** r **do**

σ_c^2, β_c, $\boldsymbol{\theta}_c \leftarrow$ Train $\mathcal{GP}_c$ model

end for

Output: GP models and indices of selected features: $\{\mathcal{GP}\}_{c=1}^{r}$, $\{\phi_i\}_{i=1}^{r}$

On-going work and challenges

- Simple parametric model as first guess (e.g. single scattering approximation)
- Refine it with a linear combination of spectral features
  → constructing spectral library
- Prediction accuracy limited by dimensionality reduction (i.e. model fitting)
- Optimizing runtime of this multi-fidelity approach still challenging

REFERENCES

- J. Vicent et al.: Comparative analysis of atmospheric radiative transfer models using the Atmospheric Look-up table Generator (ALG) toolbox (version 2.0), Geosci. Model Dev., 13, 1945–1957, DOI:10.5194/gmd-13-1945-2020, 2020.
- J. Vicent et al.: Systematic Assessment of MODTRAN Emulators for Atmospheric Correction, IEEE TGRS, 60, 4101917, DOI:10.1109/TGRS.2021.3071376, 2021
- J. Vicent et al., Multifidelity Gaussian Process Emulation for Atmospheric Radiative Transfer Models, IEEE TGRS, 61, 5519210, DOI: 10.1109/TGRS.2023.3300460, 2023
- J. Vicent et al., Multioutput Feature Selection for Emulation and Sensitivity Analysis,IEEE TGRS., 62, 5506411, DOI: 10.1109/TGRS.2024.3358231, 2024