remote sensing

MDPI

Article

Machine Learning to Retrieve Gap-Free Land Surface Temperature from Infrared Atmospheric Sounding Interferometer Observations

Fabio Della Rocca [1,2], Pamela Pasquariello [3], Guido Masiello [3], Carmine Serio [3] and Italia De Feis [2,*]

[1] Dipartimento di Biologia, Università di Napoli Federico II, 80126 Napoli, Italy; fabio.dellarocca@unina.it
[2] Istituto per le Applicazioni del Calcolo "Mauro Picone", CNR, 80131 Napoli, Italy
[3] Dipartimento di Ingegneria, Università della Basilicata, 85100 Potenza, Italy; pamela.pasquariello@unibas.it (P.P.); guido.masiello@unibas.it (G.M.); carmine.serio@unibas.it (C.S.)
* Correspondence: italia.defeis@cnr.it

Abstract: Retrieving LST from infrared spectral observations is challenging because it needs separation from emissivity in surface radiation emission, which is feasible only when the state of the surface–atmosphere system is known. Thanks to its high spectral resolution, the Infrared Atmospheric Sounding Interferometer (IASI) instrument onboard Metop polar-orbiting satellites is the only sensor that can simultaneously retrieve LST, the emissivity spectrum, and atmospheric composition. Still, it cannot penetrate thick cloud layers, making observations blind to surface emissions under cloudy conditions, with surface and atmospheric parameters being flagged as voids. The present paper aims to discuss a downscaling–fusion methodology to retrieve LST missing values on a spatial field retrieved from spatially scattered IASI observations to yield level 3, regularly gridded data, using as proxy data LST from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) flying on Meteosat Second Generation (MSG) platform, a geostationary instrument, and from the Advanced Very High-Resolution Radiometer (AVHRR) onboard Metop polar-orbiting satellites. We address this problem by using machine learning techniques, i.e., Gradient Boosting, Random Forest, Gaussian Process Regression, Neural Network, and Stacked Regression. We applied the methodology over the Po Valley region, a very heterogeneous area that allows addressing the trained models' robustness. Overall, the methods significantly enhanced spatial sampling, keeping errors in terms of Root Mean Square Error (RMSE) and bias (Mean Absolute Error, MAE) very low. Although we demonstrate and assess the results primarily using IASI data, the paper is also intended for applications to the IASI follow-on, that is, IASI Next Generation (IASI-NG), and much more to the Infrared Sounder (IRS), which is planned to fly this year, 2025, on the Meteosat Third Generation platform (MTG).

Keywords: land surface temperature; radiative; transfer; IASI; downscaling; machine learning

Academic Editors: Gino Dardanelli, Paolo Dabove and Günther Retscher

Received: 20 December 2024
Revised: 11 February 2025
Accepted: 13 February 2025
Published: 18 February 2025

Citation: Della Rocca, F.; Pasquariello, P.; Masiello, G.; Serio, C.; De Feis, I. Machine Learning to Retrieve Gap-Free Land Surface Temperature from Infrared Atmospheric Sounding Interferometer Observations. *Remote Sens.* **2025**, *17*, 694. https://doi.org/10.3390/rs17040694

1. Introduction

LST refers to the radiative skin temperature of the land (soil or top of the canopy) and can be described as a measure of how hot or cold the surface of the Earth would feel to the touch [1]. LST is one of the 55 Essential Climate Variables, a group of linked variables that critically contributes to the characterization of Earth's climate defined by a panel of experts from the Global Climate Observing System (GCOS). Its importance covers many

fields such as meteorology, climatology, agriculture, evapotranspiration, urban planning, and environmental monitoring [2]. We can have two LST products: in situ observations and remotely sensed products. In situ observations are primarily used for validation and calibration of satellite-based LST data, rather than as standalone products. This is because LST is highly variable depending on surface materials, vegetation cover, soil moisture, atmospheric conditions, and topography, which can make in situ measurements less representative on larger scales. Thermal infrared (TIR) or Microwave (MW) remote sensing is considered to be the most efficient way to obtain accurate LST, both regionally and globally. However, as in all remotely sensed products, there is a trade-off between spatial, spectral, and temporal resolution. For example, TIR images from Moderate Resolution Imaging Spectroradiometer (MODIS), have a high temporal resolution (twice a day) but a coarse spatial resolution (1 km). Other sensors such as the Thermal Infrared Sensor (TIRS) onboard LANDSAT platforms, have a greater spatial resolution (100 m in this case) but much lower temporal resolution (8 days). In contrast, TIR images from geostationary satellites, like from SEVIRI, have a much lower revisit time (minutes) but at the cost of low spatial resolution (5 km). Having a coarse resolution, results in the thermal mixture effect, i.e., when a given thermal pixel contains a blend of multiple thermal signals that are smaller than the resolution cell [3]. Therefore, there is a growing need to improve the spatial resolutions of LST products.

The retrieval in the LST commonly relies on one or more channels within the TIR region, spanning 8–13 μm. This interval, known as the infrared atmospheric window, is a portion of the infrared spectrum where atmospheric gases absorb relatively little terrestrial thermal radiation, allowing the atmosphere to be more transparent. This window plays a crucial role in the atmospheric greenhouse effect by maintaining the balance between incoming solar radiation and outgoing infrared radiation to space. Satellite sensors take advantage of this window for the retrieval of an LST with minimal atmospheric distortion. These channels also correspond to the wavelengths of the peak monochromatic radiance, predicted by Wien's displacement law, for typical LST values (ranging from approximately 250 K to 330 K). The retrieval of the LST from TIR radiances is a physical process that involves the inversion of the Radiative Transfer Equation (RTE). However, estimating LST from satellite-measured radiance requires corrections for both atmospheric and emissivity effects, which is a challenging task. The inversion of the RTE is an underdetermined and ill-posed problem [4,5], even when atmospheric parameters are known, for N channels there are still N + 1 unknowns (N emissivities and the LST, which is not wavelength related). To make the retrieval of LST a deterministic problem, it is necessary to know the value of at least one emissivity in advance. Alternatively, the LST and Land Surface Emissivities (LSEs) can be estimated simultaneously by imposing physical assumptions or constraints on the LSE [2,4–7]. Numerous algorithms have been proposed for estimating LST from satellite observations, which can be broadly classified into three categories: single-channel algorithms, multi-channel methods, and multi-temporal methods [8]. The single-channel method [9–11] is the simplest approach, relying on the inversion of the radiative transfer equation, provided that Land Surface Emissivities (LSEs) and atmospheric profiles are known in advance. Multi-channel methods include the Split-Window (SW) algorithm and the Temperature and Emissivity Separation (TES) method. Among these, the SW algorithm is the most commonly used for LST retrieval from satellite data and takes advantage of the differential atmospheric absorption in two adjacent TIR channels [12]. Examples of LST products generated using an SW algorithm include MODIS LST (MOD11) from the National Aeronautics and Space Administration (NASA) [13], VIIRS (Visible Infrared Imaging Radiometer Suite) LST from the National Oceanic and Atmospheric Administration (NOAA) [14], Sentinel-3 LST from the Copernicus Sea and Land Surface Temperature Ra-

diometer (SLSTR) [15], GOES-16 LST from the Advanced Baseline Imager (ABI) onboard NOAA's Geostationary Operational Environmental Satellites (GOES) [16], and LST products from the LSA SAF (Satellite Application Facility on Land Surface Analysis) derived from SEVIRI and AVHRR sensors [17]. However, this method still requires the LSEs to be known a priori, and LST retrieval accuracy is degraded under a highly humid atmosphere or at large viewing zenith angles [8,18–20]. An improvement over the SW algorithm is provided by the TES algorithm [7], which utilizes multiple TIR channels and an empirical relationship between the minimum LSE and the difference between maximum and minimum LSE to dynamically retrieve both LST and emissivity. An example of its application is NASA's LST MOD21 from MODIS [21]. However, the TES algorithm's performance is limited by the accuracy of the atmospheric correction, and the results are inaccurate for surfaces with low spectral emissivity contrast (i.e., similar emissivity in different channels), such as dense vegetation, water, and snow [8]. Multi-time methods combine multi-temporal information under the assumption that LSEs do not change significantly between adjacent day and night observations. For example, the physical day and night algorithm [22] employs both TIR and mid-infrared (MIR) bands to accurately retrieve LSTs and LSEs in a physical way. However, the retrieval is complicated by the fact that two clear, independent, and perfectly co-registered observations are needed. Additionally, it still requires an a priori approximate atmospheric profile, and it assumes constant emissivity throughout the day, which may not be true (e.g., due to condensation or dew).

Due to its high spectral resolution, the IASI instrument onboard Metop polar-orbiting satellites is the only sensor capable of simultaneously retrieving LST, the emissivity spectrum, and atmospheric composition. This characteristic addresses the limitations of previous methods, as it permits improvement over a prori information on emissivity, temperatures, and gases. IASI has the advantage of having an excellent spectral and temporal resolution, with a spatial resolution of 12 km at nadir. These characteristics allowed us to design a scheme for LST retrieval (e.g., see [23,24]) that considers many factors that govern LST. Our scheme simultaneously retrieves LST and the whole emissivity spectrum. In addition, we also consider the viewing angle, air temperature, and water vapor mixing ratio, which makes our methodology unique in its capability of capturing LST variability and its day-night variations.

Since it is an infrared sensor, it cannot penetrate thick cloud layers and observations are blinded to surface emissions due to cloudiness, bringing sparse and non-homogeneous distributed data over a given spatial region (L2 observations). This requires robust techniques to fill the gaps and retrieve a gridded product.

In effect, the present study is motivated by the need to define and validate suitable retrieval methods, including those for LST, for the IASI Next Generation (IASI-NG) instrument (e.g., [25]), and the Infrared Sounder (IRS). The IASI-NG instrument is planned to launch in 2025, continue the IASI mission, and is characterized by a spectral resolution that is twice better than that of IASI. IASI-NG is a joint venture of Centre National d'Etudes Spatiales (CNES)-EUMETSAT (built at CNES, Toulouse, France, see https://cnes.fr/en/projects/iasi-ng (accessed on 12 February 2025)) and will fly on board the Metop Second Generation (SG) satellite. The complex retrieval algorithm we developed for IASI can be extended to IASI-NG in a straightforward manner. IRS is a Fourier Transform Infrared Spectrometer, the same class as IASI, and has been designed to fly on the Meteosat Third Generation satellite in 2025. IRS is a joint venture of European Space Agency (ESA)-EUMETSAT (built at Thales Alenia Space, Cannes, France and OHB, Bremen, Germany see https://www.eoportal.org/satellite-missions/meteosat-third-generation#performance-specifications (accessed on 12 February 2025)) , and the instrument has a spectral resolution that is half that of IASI. The spectral coverage has some limitations com-

pared to IASI and is composed of two bands, one looking at the longwave spectral region (700 to 1210 cm^{-1}) and the other one at the mid-infrared (1600 to 2175 cm^{-1}). However, IRS can fly on a geostationary platform and can follow surface and atmospheric processes, allowing us to extend and improve forward/inverse methods we previously developed, e.g., for SEVIRI ([26–28]).

Many studies have focused on interpolation and downscaling of LST products to produce high-resolution LST data [29], with an increasing application of Machine Learning (ML) techniques. The biggest advantage of ML algorithms is that they are flexible and not restricted to linear relationships (as in linear regression or regression kriging). This flexibility enables the production of accurate results in modelling the relationship between LST and covariates. For instance, Refs. [30–32] demonstrated that the Random Forest (RF) algorithm can perform spatial interpolation which is on par with or even surpasses traditional state-of-the-art techniques. Since LST is an essential environmental parameter, researchers have made many efforts to improve its resolution. Given the heterogeneity of LST and its relationship to several types of environmental factors such as land cover, topography, soil moisture, incoming solar radiation, and wind speed [2,33], ML algorithms are becoming increasingly popular due to their ability to model this nonlinear relationship. For example, Refs. [34–40] proved that a simple RF model can outperform many other ML and NN algorithms in downscaling LST images from MODIS. For example, in [35], the authors reached a final resolution ten times more detailed than the original one (from 990 to 90 m) with an RMSE of around 2 K. Deep Learning is also involved in LST downscaling, like with Convolutional Neural Networks (CNN) [41], geographically neural network weighted regression (GNNWR) [42], and Multi-layer Perceptron [43,44].

However, almost all above-cited studies focus on downscaling LST images from MODIS, see also [45] and from other sensors, such as AVHRR [46,47], ASTER [48,49], LANDSAT [50–55], SEVIRI [56,57], or GOES [58]. See [29] and references therein.

To the best of our knowledge, the downscaling of LST products generated from IASI has not yet been addressed. Some efforts have been made to retrieve regular grid LST using Deep Neural Networks. For example, Ref. [59] shows an LST product obtained via the application of a NN to radiances. However, to improve the results, they incorporated a monthly surface emissivity product, and the predicted LST were subsequently interpolated. Other examples are described in [60–62] where localized CNNs were tested to predict LST directly from radiances. Since CNNs operate on images, the authors converted the L1 products into regular grid images at 0.2 degrees resolution. However, missing data are still present making an inter-/extrapolation algorithm based on PCA necessary to facilitate smooth training.

In this context, we applied a double-step procedure in order to retrieve an IASI L3 LST product. First, we considered the physical inversion of the RTE using the ϕ-IASI package, a FORTRAN-based inverse radiative transfer code for the optimal estimation of the thermodynamic state of the atmosphere. It consists of two modules, a forward model called σ-IASI [63,64] and an inverse one called δ-IASI [23,65–67]. Ref. [67] refers to the newest version of the code available upon request. Due to IASI's outstanding spectral resolution, our approach retrieves LST, LSE, and atmospheric composition simultaneously starting from a priori information derived from both climatological, gases, and emissivity databases. Second, we exploited the capability of different ML algorithms in converting LST IASI L2 scattered data to a regular L3 grid LST, increasing also the spatial resolution of the final product. Specifically, we trained a model predicting the LST index over the L2 observations using covariate data from other sensors such as vegetational products, soil indices, and territorial and geographic information. These data are colocalized temporally and spatially onto the L2 observations. Afterwards, downscaling of LST is performed using

the higher resolution of the covariates to predict the index over a 0.05° regular grid, actually doubling the spatial resolution.

We applied the methodology over the Po Valley region, which experienced an intense drought in the last three years, which caused high vegetation and soil water stress. Moreover, since the Po Valley region is a very heterogeneous area, we can also address the ability of the trained models to generalize and give a good prediction over different area types.

The structure of the paper is as follows: Section 2 illustrates the region of interest, the data we used, and the description of the methods; Section 3 presents the results and evaluates the LST downscaling result. Finally, Section 4 discusses the results and Section 5 draws conclusions.

2. Materials and Methods

2.1. Study Area

Our study area is the Po Valley (44°–46°N, 7°–13°E WGS84) located in northern Italy and covers an area of approximately 105,096 km^2. This large area contains the Alps, the Apennines, and the Po Valley showing strong differences in altitudes, vegetation cover, and climate. Figure 1 illustrates the region of interest, with land cover information based on the CLC 2018 dataset, available at https://land.copernicus.eu/pan-european/corine-land-cover (accessed on 12 February 2024).

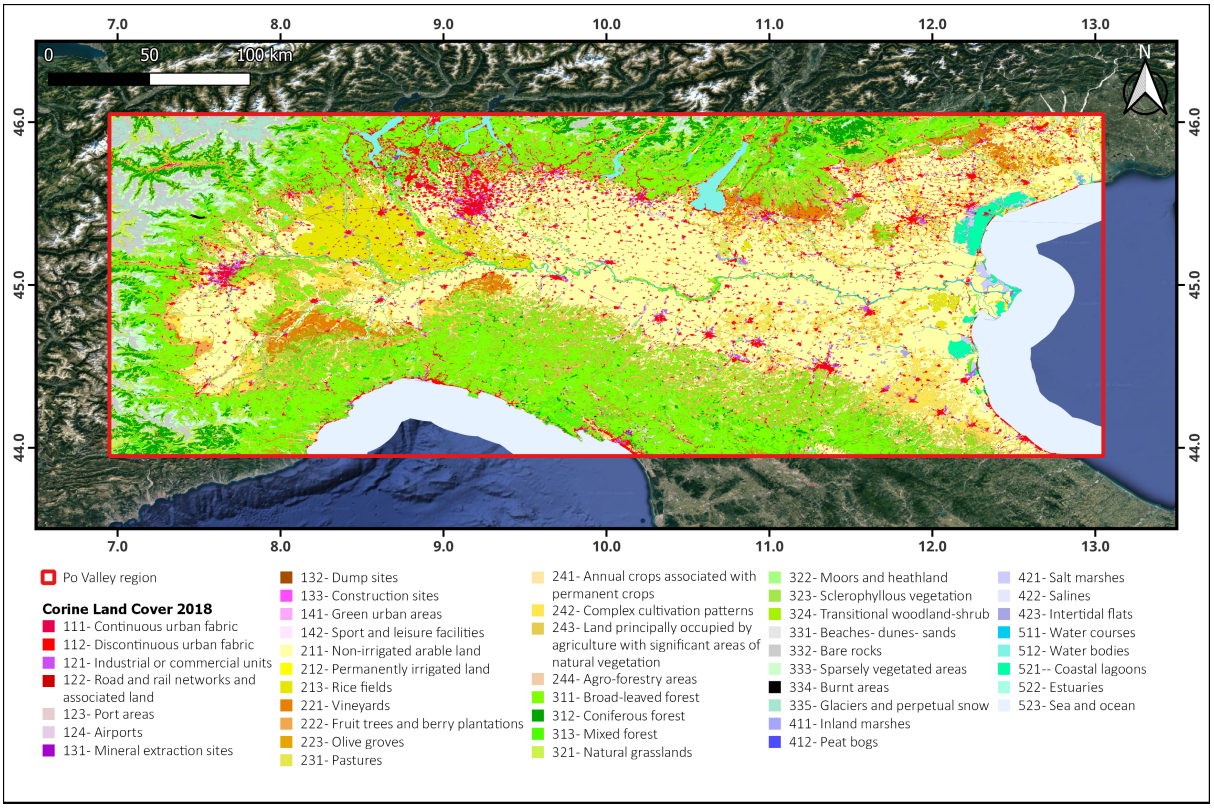

Figure 1. The red box indicates the Po Valley target region with the CLC 2018 as shapefile.

The Po Valley region is a flat alluvional plain crossed by the Po River and highly devoted to agriculture. The plain is surrounded by the Apennines in the southern part and by the Alps in the northern and western parts. The Apennines, which rarely exceed 2000 m in altitude, extend 1200 km along the Italian peninsula, hosting a rich variety of flora and fauna. These areas contain some of Europe's best-preserved montane grasslands and forests, mainly composed of broad-leaved woodlands. In contrast, the Alps form

Europe's largest and highest mountain range, with many peaks exceeding 4000 m and covered by broad-leaved and coniferous forests.

2.2. Data

2.2.1. IASI Data

The LST we used was retrieved from the IASI sensor [68], a Fourier transform spectrometer that measures infrared radiation emitted from the Earth. IASI was built by the CNES and flew on board the Metop platforms, satellites of the EUMETSAT European Polar System (EPS). The spectral coverage of the instrument extends from 645 to 2760 cm^{-1} with a spectral sampling of 0.25 cm^{-1} providing a total of 8461 channels. IASI is a cross-track scanner with 30 adjacent fields of regard (FOR) per scan. Each FOR covers a square area of 50 × 50 km^2 on the ground and consists of a 2 × 2 matrix of instantaneous fields of views (IFOVs) with a diameter of 0.8394° (corresponding to a ground resolution of 12 km at nadir for a satellite altitude of 819 km). We downloaded 9 years (from 2015 to 2023) of IASI data (Level 1C) from the Eumetsat Data Centre at https://archive.eumetsat.int/usc/UserServicesClient.html (accessed on on 21 March 2024), for a total of almost 10,000 orbits coming from MetOp-A, MetOp-B, and MetOp-C platforms (here we considered only day passages).

2.2.2. Geographical Covariates

The model we aim to build should perform spatial regression, meaning it should predict new values while considering the spatial dependence of observations. In methods like Kriging, spatial dependence is integrated within the covariance matrix, but this type of information is harder to incorporate in an ML model. A straightforward approach is to include spatial coordinates (latitude and longitude) directly as input features. Additionally, we used the European Digital Elevation Model (DEM) and the Corine Land Cover (CLC) 2018 to better capture territorial characteristics. The DEM https://land.copernicus.eu/imagery-in-situ/eu-dem (accessed on 1 December 2023) is a raster with a 25 m spatial resolution, while the CLC https://land.copernicus.eu/pan-european/corine-land-cover (accessed on 12 February 2024) is a shapefile. Following a method similar to [69], we converted the CLC into a raster format with 100 m resolution, simplifying it to five main classes corresponding to the Level 1 CLC classification: artificial surfaces, agricultural areas, forest and seminatural areas, wetlands, and water bodies. A more detailed classification reduced model performance because spatial co-localization becomes challenging when assigning a single detailed CLC class to pixels with a minimum diameter of 12 km. We are aware that Copernicus provides also a CLC in a raster format at 100 m, but this data is created by the rasterization of seamless European CLC vector data using the CELL_CENTER method https://desktop.arcgis.com/en/arcmap/latest/tools/conversion-toolbox/how-polygon-to-raster-works.htm (accessed on 15 February 2024). LST varies significantly with different land cover types [70,71] because each of them has distinct thermal properties that influence heat absorption and emission. The LST provided by IASI refers to a huge area of 12 km in diameter, hence averaging the signals from a broad variety of land cover types. The CELL_CENTER method assigns to a given pixel the CLC class that exactly falls in the centre. Our approach, instead, considers the CLC class with the highest area.

2.2.3. Vegetational Covariates

It is well known that healthy vegetation mitigates the LST and the air temperature by modifying albedo, evapotranspiration, and providing shade [72–75]. The evapotranspiration converts solar energy into latent heat, reducing the air temperature and, as a consequence, the LST. A high canopy density of vegetation communities results not only in a greater shadowing effect (lower air temperatures and LST) but also in greater water

retention under the canopy, that is higher air humidity and a higher intensity of humidification, determining a better cooling effect [76]. There is, hence, a clear relationship between vegetation indexes and LST, which is helpful in improving the downscaling process, as in [77]. We used, as covariates, three different vegetation indices: Leaf Area Index (LAI), Fractional Vegetation Cover (FVC), and Normalized Difference Vegetation Index (NDVI). LAI quantifies the thickness of the vegetation cover and is defined as half the total area of green elements of the canopy per unit of horizontal ground area. FVC corresponds to the fraction of ground covered by green vegetation and thus, quantifies the spatial extent of the vegetation. NDVI is a widely used index for quantifying the health and density of vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). We used the EPS LSA SAF products, available at the following link https://landsaf.ipma.pt/en/data/products/vegetation/ (accessed on 29 April 2024), with a temporal resolution of 10 days and a spatial resolution of 1 km (corresponding to 0.0089°). The products are derived from AVHRR, an across-track scanner carried by the MetOp platforms (the same as IASI). We used only the good quality pixels, as suggested by the LSA SAF in the Product User Manual (https://nextcloud.lsasvcs.ipma.pt/s/sWn23j9NZnJg2a6?dir=undefined&path=%2FPUM&openfile=27136 (accessed on 29 April 2024)). Specifically, we discarded observations with poor quality (bit 7 = 1), large estimating errors (FVC_ERR > 0.2, LAI_ERR >1.5), and large age of observation (Z_Age >15, associated to snow or persistent cloudiness).

2.2.4. Other LST Products

To further improve the quality of the final prediction we also included two LST products: an LST derived from the AVHRR sensor and an LST derived from the SEVIRI sensor. We downloaded the two LST products from LSA SAF at https://lsa-saf.eumetsat.int/en/data/products/land-surface-temperature-and-emissivity/ (accessed on 6 May 2024). The temporal and spatial resolution of the two products are respectively 15 min and 0.05° for LST SEVIRI, and 12 h (day and night) and 0.01° for LST AVHRR. Only pixels with nominal or above-nominal quality (LST uncertainties below 2 K) were considered. This uncertainty threshold represents a compromise, ensuring high-quality LST values while preserving a sufficient number of data points for the validation.

For comparison, we used the LST from MODIS, specifically the MOD11A1 Version 6.1 dataset from NASA https://lpdaac.usgs.gov/products/mod11a1v061/ (accessed on 8 October 2024). MODIS LST is a widely used and extensively validated product [78,79], making it an ideal candidate for comparison. The MODIS sensors, carried on the Terra and Aqua satellites, have provided daily data since the years 2000 (Terra) and 2002 (Aqua) with a spatial resolution of about 1 km at the nadir in the TIR channels (31 and 32) used for the computation of the LST. The MOD11 version used in this study is derived exclusively from the Terra satellite, which has acquisition times closest to IASI, and is processed using the generalized split-window algorithm [13]. It is well known that in semi-arid and arid areas, there may be significant uncertainty in the estimated emissivities of the MOD11 product, but this should not be the case of the Po Valley region. Furthermore, the MOD11 product appears to be more stable compared to the MOD21, which has more extreme values, possibly due to incorrect emissivity retrieval or undetected clouds [80]. Again, only MODIS LST data with good quality and an uncertainty smaller than 2 K were retained for comparison.

2.3. IASI Retrieval

The physical inversion of IASI was performed using the ϕ-IASI package, a FORTRAN-based inverse radiative transfer code for the optimal estimation of the thermodynamic

state of the atmosphere. This package, which was recently updated to match the spectral coverage of the Far-infrared Outgoing Radiation Understanding Monitoring (FORUM) mission [67], consists of two modules, the forward model, σ-IASI [63,64] and the inverse one, δ-IASI [23,24,65,66]. The forward model is a pseudo-monochromatic, fast line-by-line radiative transfer scheme which computes the surface/atmosphere spectrum in the infrared and related Jacobian matrices. The core of σ-IASI is the following radiative transfer equation valid for computing a clear-sky radiance:

$$\begin{aligned} R(\sigma) &= R^{surf}(\sigma) + R^{\uparrow}(\sigma) + R^{\downarrow}(\sigma) + R^{\odot}(\sigma) = \\ &= \varepsilon_g(\sigma) B(\sigma, T_g) \tau_0(\sigma) + \int_0^{+\infty} B(\sigma, T(z)) \frac{\partial \tau(\sigma)}{\partial z} dz \\ &+ (\varepsilon_g(\sigma) - 1) \tau_0(\sigma) \int_0^{+\infty} B(\sigma, T(z)) \frac{\partial \tau_*^f(\sigma)}{\partial z} dz + \\ &+ \frac{1 - \varepsilon_g(\sigma)}{\pi} \tau_0(\theta_{sat}) \tau_0(\theta_{sun}) \mu_{sun} I_{sun}(\sigma) \end{aligned} \tag{1}$$

where

- $\varepsilon_g(\sigma)$ is the surface emissivity;
- T_g is the skin temperature;
- $\tau_0(\sigma)$ is the total transmittance of the atmosphere (along a vertical path);
- $B(\sigma, T) = \frac{c_1 \sigma^3}{\exp(c_2 \sigma / T) - 1}$ is the Planck blackbody source term with $c_1 = 1.1911 \times 10^{-8}$ $\mathrm{Wm^{-2}sr^{-1}}\left(\mathrm{cm^{-1}}\right)^{-4}$ and $c_2 = 1.4388\ \mathrm{K}\left(\mathrm{cm^{-1}}\right)^{-1}$;
- $\tau(\sigma)$ is the transmittance along $ds = dz/\mu$ $\mu = \cos(\theta)$ (with θ the satellite zenith angle) from altitude z to $+\infty$;
- $\tau_*^f(\sigma) = 2\int_0^1 \tau_*(\mu, z)\mu d\mu$ and τ_* is the transmittance from altitude z to $z = 0$ along the slant path in the direction μ;
- $I_{sun}(\sigma)$ is the solar spectrum [81] incident on a normal surface and normalized to the solar disk solid angle as seen by the Earth (6.8×10^{-5} sr) [W m^{-2} (cm^{-1})$^{-1}$], multiplied by $\mu_{sun} = \cos(\theta_{sun})$;
- $\tau_0(\theta_{sun})$ is the total transmittance along the downwelling θ_{sun}-direction;
- $\tau_0(\theta_{sat})$ is the total transmittance along the upwelling θ_{sat}-direction.

It is important to stress that the top-to-bottom transmittance τ_* should not be confused with the bottom-to-top transmittance τ. According to W.M. Elsasser [82] it is postulated that

$$\tau_*^f = \tau_*(\bar{\mu}, z), \tag{2}$$

i.e., the diffuse transmittance can be calculated as the transmittance function at a suitable cosine angle, the term $\frac{1}{\bar{\mu}}$ is referred to as the diffusivity factor and for practical calculations, the value $\frac{1}{\bar{\mu}} = 1.66$ (corresponding to an effective zenith angle of 52.96° yields accurate results, which depend on the optical depth; the effective zenith angle is spectrally dependent. For optical depth below 1, exact and approximate (through Equation (2)) calculations nearly coincide. Equation (1) applies to the case of a non-scattering atmosphere where only the absorption by atmospheric gas is considered and is a good approximation for a clear sky. In presence of clouds and aerosols, a source function is needed to properly deal with multiple scattering, see [67]. The downward infrared radiation reflected at the surface $R^{\downarrow}(\sigma)$ is fundamental for emissivity retrieval, because it depends on emissivity alone, while the surface term $R^{surf}(\sigma)$ depends on the product of emissivity and the Planck function computed at the surface temperature.

For numerical calculations, we need to compute the atmospheric emission integrals in Equation (1) for the upwelling and downwelling contribution. To this end, considering L

atmospheric layers, i.e., the grid $z_0, z_1, \cdots, z_L$, and adopting the approximation of parallel-plane atmosphere, Equation (1) becomes the following:

$$R(\sigma) = \varepsilon_g B(T_g)\tau_0 + \sum_{j=1}^{L} B(T(z^*))(\tau_j - \tau_{j-1}) + (\varepsilon_g - 1)\tau_0 \sum_{j=1}^{L} B(T(z^*))(\tau_j^* - \tau_{j-1}^*) + \\ + \frac{1-\varepsilon_g(\sigma)}{\pi}\tau_0(\theta_{sat})\tau_0(\theta_{sun})\mu_{sun} I_{sun}(\sigma) \tag{3}$$

with $z_{j-1} \leq z^* \leq z_j$.

The retrieval module, δ-IASI, implements an iterative algorithm for the optimal estimation of the thermodynamic state of the atmosphere. The retrieval algorithm follows Rodgers optimal estimation method [83] by considering an iterative Gauss–Newton scheme. In details, let us indicate with $\mathbf{R}_{obs} = (R(\sigma_1), \cdots, R(\sigma_N))^T \in \mathbb{R}^{N\times 1}$ the vector of observed radiances, with $\mathbf{v} = (v_1, \cdots, v_M)^T \in \mathbb{R}^{M\times 1}$ the vector of geophysical parameters (atmospheric and surface), then we can express Equation (3) in compact form as

$$\mathbf{R}_{obs} = F(\mathbf{v}) + \boldsymbol{\epsilon} \tag{4}$$

where $\boldsymbol{\epsilon}$ is the additive noise which accounts for observational, i.e., instrumental detector noise $\boldsymbol{\epsilon}_r$, and forward model bias $\boldsymbol{\epsilon}_m$. We make the hypothesis that the model noise $\boldsymbol{\epsilon}_m$ can be neglected with respect to the random error $\boldsymbol{\epsilon}_r$, that we assume Gaussian with $\mathbf{0}$ mean and covariance $\mathbf{S}_\epsilon$ (we omitted the subscript r to simplify notation). Assuming Gaussian a priori information about the state parameter $\mathbf{v}$, independent from the instrumental one, i.e.,

$$p(\mathbf{v}) \sim N\left(\mathbf{v}_a, \frac{1}{\gamma}\mathbf{S}_a\right), \tag{5}$$

where $\gamma \in (0, +\infty)$ quantifies how much we predict that $\mathbf{v}$ should be close to $\mathbf{v}_a$, then the predictive distribution $[\mathbf{v}|\mathbf{R}_{obs}]$ is no longer Gaussian due to the nonlinearity of F

$$-2\ln P(\mathbf{v}|\mathbf{R}_{obs}) = (\mathbf{R}_{obs} - F(\mathbf{v}))^T \mathbf{S}_\varepsilon^{-1}(\mathbf{R}_{obs} - F(\mathbf{v})) + \gamma(\mathbf{v} - \mathbf{v}_a)^T \mathbf{S}_a^{-1}(\mathbf{v} - \mathbf{v}_a). \tag{6}$$

The maximum probability state $\hat{\mathbf{v}}$ is the solution to the minimization problem

$$\underset{v}{\operatorname{argmin}}(\mathbf{R}_{obs} - F(\mathbf{v}))^T \mathbf{S}_\varepsilon^{-1}(\mathbf{R}_{obs} - F(\mathbf{v})) + \gamma(\mathbf{v} - \mathbf{v}_a)^T \mathbf{S}_a^{-1}(\mathbf{v} - \mathbf{v}_a), \tag{7}$$

and relaxing to the classical Gauss–Newton iteration scheme, we obtain

$$\mathbf{v}_{\ell+1} = \mathbf{v}_a + \left(\gamma_\ell \mathbf{S}_a^{-1} + \mathbf{K}_\ell^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_\ell\right)^{-1} \mathbf{K}_\ell^T \mathbf{S}_\varepsilon^{-1}[(\mathbf{R}_{obs} - F(\mathbf{v}_\ell)) + \mathbf{K}_\ell(\mathbf{v}_\ell - \mathbf{v}_a)], \tag{8}$$

where $K_\ell = \frac{\partial F}{\partial \mathbf{v}}|_{\mathbf{v}=\mathbf{v}_\ell}$ is the jacobian and a typical choice for the starting value $\mathbf{v}_0$ for $\ell = 0$ (*first guess*) in the iteration is $\mathbf{v}_a$.

At the starting iteration $\ell = 0$, because the derivative is additive, the part related to the emissivity in the state vector and that to the other parameters can be separated,

$$\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2) \quad \text{and} \quad \mathbf{v} - \mathbf{v}_0 = \begin{pmatrix} \mathbf{v}_1 - \mathbf{v}_{01} \\ \mathbf{v}_2 - \mathbf{v}_{02} \end{pmatrix}$$

where the subscripts 1 and 2 refer to the atmospheric and emissivity components, respectively. The number of atmospheric parameters will be denoted by M_1, while that corresponding to emissivity by M_2 (note that $M_2 = N$ number of radiance data, i.e., spec-

tral channels). The Jacobian $\mathbf{K}_1$ is computed as usual as the derivative of the radiance, $R(\sigma)$, with respect to the surface temperature and atmospheric parameters. In the same way, $\mathbf{K}_2$ is computed by differentiating $R(\sigma)$ with respect to emissivity. However, the computation of $\mathbf{K}_2$ is less straightforward because we first transform emissivity to project it into an unbounded space (spanning from $-\infty$ to $+\infty$) and then reduce its dimensionality.

We achieve this by transforming emissivity with the logit function,

$$z(i) = \text{logit}((\varepsilon(i))) = \log\left(\frac{\varepsilon(i)}{1-\varepsilon(i)}\right) \Rightarrow \varepsilon(i) = \frac{\exp(z(i))}{1+\exp(z(i))} \quad i = 1, \ldots, M_2$$

The logit transform avoids the retrieval of un-physical emissivities and/or the use of boundary constraints on each parameter because it allows us to work with a quantity, z, which is defined in the range $[-\infty, +\infty]$. Once back transformed, the retrieved emissivity is forced to be a number in the interval $[0, 1]$.

The jacobian can be easily derived using the chain rule

$$\frac{\partial R(i)}{\partial z(i)} = \frac{\partial R(i)}{\partial \varepsilon(i)}\left(\frac{\partial z(i)}{\partial \varepsilon(i)}\right)^{-1} = \frac{\partial R(i)}{\partial \varepsilon(i)}\varepsilon(i)(1-\varepsilon(i)), \quad i = 1, \ldots, M_2$$

and the component pertinent to emissivity reads

$$\mathbf{K}_{\mathbf{z},z}(\mathbf{z} - \mathbf{z}_0)$$

with $\mathbf{K}_{\mathbf{z},2}$ a diagonal matrix with elements $K_{\mathbf{z},2}(i,i) = \frac{\partial R(i)}{\partial z(i)}$.

Second, we develop the z spectrum in a truncated PCA series, with truncation point, $\tau \leq M_2$. To develop the PCA basis we used ASTER (Advanced Spaceborne Thermal Emission Reflection Radiometer) Spectral Library version 2.0 [84] and the MODIS (Moderate Resolution Imaging Spectrometer) UCSB (University of California, Santa Barbara) Emissivity Library (http://www.icess.ucsb.edu/modis/EMIS/html/em.html (accessed on 22 March 2024)). The ensemble counts 134 representative emissivity spectra at a global scale, of senescent and green vegetation, bare soil, desert sand, and rock emissivities. The emissivity spectra are laboratory measurements at a spectral sampling of 2 cm^{-1}. They have been linearly re-sampled at the IASI sampling of 0.25 cm^{-1} and re-interpolated to the IASI range 645–2760 cm^{-1}. Generally, 20 PC scores are sufficient to represent all the important spectral features. Indeed, the emissivity of natural and terrestrial materials in the infrared can be fully resolved at a spectral resolution of 2 cm^{-1} (e.g., [84]); therefore, 20 PCA scores should be enough to correctly represent emissivity of Earth's surface. See [24,85] for more details and retrieval over heterogeneous surfaces.

By inverting the full IASI spectrum (8461 spectral channels), δ-IASI retrieves all the major atmospheric components such as surface emissivity (ϵ), surface temperature (Ts), vertical profiles of temperature (T), water vapour (Q), ozone (O), HDO mixing ratio (D) and average column abundances of CO_2, CO, CH_4, N_2O, HNO_3, SO_2, NH_3, OCS and CF_4 [23,24,85–87].

The ϕ-IASI algorithm automatically discards cloud pixels using the internal AVHRR cloud mask: pixels with a cloud fraction greater than 5% are removed. This introduces missing data, requiring an interpolation algorithm to transform the sparsely populated, non-uniform L2 observations (ϕ-IASI output), into evenly-gridded maps. Ref. [88] developed a procedure similar to Gaussian Optimal Interpolation with the inclusion of background data, which converted the L2 observations into monthly maps with a resolution of 0.05°. However, this procedure, which was already tested on many surface and atmospheric parameters [69,89–91], generates maps that are quite smoothed. Although this approach is perfectly suitable for parameters like air temperature or gases, which typically show little variation over short distances, it does not hold true for LST, which can vary sig-

nificantly. This study aims to address this issue, which will be further discussed in the subsequent chapters.

2.4. Spatial and Temporal Colocalization

We spatially and temporally collocated all the downloaded and preprocessed data with the IASI L2 observations to construct the L2 training dataset. As stressed in Section 2.2.1, each IASI observation has a spatial resolution of 12 km at the nadir, which can extend to more than 20 km when observations are off-nadir. Additionally, the actual shape of the IASI footprint is circular only at nadir, becoming elongated and elliptical toward the edges of the swath. Thus, accurately considering each IASI observation's footprint is necessary. The IASI FOV angle ("pixel_zenith_angle") alone does not suffice to determine the footprint shape, as the ellipse's major and minor axes can be oriented in any direction. This requires knowledge of the IASI orbit, adding complexity to the colocalization process. To simplify, we approximate the elliptical footprint as a circle with a diameter equal to the ellipse's minor axis. This conservative approximation ensures that the entire circular area falls within the actual footprint. For spatial colocalization, since the covariates have a finer spatial resolution than IASI, we averaged the covariate values within each IASI footprint. Since the CLC is categorical, we assigned the most frequent class within each footprint.

More attention should be paid to temporal colocalization since the chosen covariates have different temporal resolutions. For the vegetational products (LAI, FVC, NDVI) and LST AVHRR, which have respectively ten days and 12 h temporal resolution, we took the closest observation considering the IASI acquisition day. LST SEVIRI, on the other hand, has a temporal resolution of just 15 min; therefore, we selected the closest observation, taking into account the exact IASI acquisition time.

The so-built training dataset is cleaned of missing values and outliers. Outliers are detected considering the differences between the IASI LST values and the corresponding values from two other LST products. This filtering is crucial, as we found pixels with differences exceeding 10 °C, which could negatively impact the performance of the trained models. Specifically, for each month, we applied the Interquartile Range (IQR) Method to the linear differences between the LSTs rather than to the LST values themselves. The outliers are defined as any values that fall outside the range (25th Percentile $-$ 3 $\times$ IQR) to (75th Percentile + 3 $\times$ IQR), where IQR = 75th Percentile $-$ 25th Percentile. We selected an IQR threshold of 3 to focus on removing only the most extreme outliers while preserving the majority of observations. This choice strikes a balance between filtering out major anomalies and retaining valuable data for model training.

2.5. ML Algorithms

We used the built dataset to train a model to predict the LST over a regular grid. We considered different ML algorithms, i.e., Random Forest (RF) [92], Gradient-Boosted Trees (GBT) [93], Neural Network (NN) [94], Gaussian Process Regression (GPR) [95], and combined them using Stacked Regression [96]. From a mathematical point of view, we modeled the L2 spatial observation y_i, $i = 1, \cdots, n$ as $y_i = \mu(\mathbf{x}_i) + \varepsilon_i$, where μ is the regression function to be estimated and $\mathbf{x} \in \mathbb{R}^d$ is the features vector.

- RF and GBT are both ensemble methods that combine multiple decision trees, such as Decision Trees or Classification and Regression Trees (CART), to improve predictive performance.
  In Random Forests (RF), M trees are created through bagging [92] and random feature selection. Each tree is independently trained on a randomly selected subset of the training data, with a series of splitting rules applied to build the tree. The essential idea in bagging is to average many noisy but approximately unbiased models, and reduce

the variance. Trees are ideal candidates for bagging, since they can capture complex interaction structures in the data, and if grown sufficiently deep, have relatively low bias. See [97] for more details.

The bagging process can be summarized as follows:

1. For $m = 1, \ldots, M$ (where M is the total number of trees):
   (a) Sample, with replacement, k training examples from $\mathbf{X}, \mathbf{Y}$; call these $\mathbf{X}_m, \mathbf{Y}_m$.
   (b) Train a classification or regression tree f_m on $\mathbf{X}_m, \mathbf{Y}_m$ considering only a random subset of the features. This random feature selection enhances decorrelation between trees. Recursively repeat the following steps for each terminal node of the tree, until the minimum node size is reached.
      i. Select p variables at random from the d variables.
      ii. Pick the best variable/split-point among the p.
      iii. Split the node into two daughter nodes.
2. After training, predictions for unseen samples $\mathbf{x}'$ are made by aggregating the predictions from all individual trees, and for regression tasks, the final prediction is obtained by averaging:

$$\hat{f}(\mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} f_m(\mathbf{x}').$$

By aggregating M decorrelated trees, RF effectively reduces variance and produces a robust final estimate.

- Gradient Boosted Trees (GBT), on the other hand, use boosting [93] to aggregate trees. The key difference is that in bagging, each tree is built independently, while in boosting, each new tree is added sequentially, specifically designed to correct errors made by previous trees.

  A generic GBT algorithm for regression, as described in [97], can be written as follows:

  1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$ where L is the loss function.
  2. For $m = 1$ to M:
     (a) For $i = 1, \ldots, n$, compute the pseudo residuals:

$$r_i^{(m)} = -\left[\frac{\partial L(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)}\right]_{f=f_{m-1}}.$$

     (b) Fit a regression tree to the targets $r_i^{(m)}$, giving terminal regions R_{jm}, $j = 1, \ldots, J_m$.
     (c) For $j = 1, \ldots, J_m$ compute:

$$\gamma_{jm} = \arg\min_\gamma \sum_{\mathbf{x}_i \in R_{jm}} L\big(y_i, f_{m-1}(\mathbf{x}_i) + \gamma\big).$$

     (d) Update:

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \eta \sum_{j=1}^{J_m} \gamma_{jm} \mathbb{I}(\mathbf{x} \in R_{jm}),$$

     where $\mathbb{I}(\cdot)$ is the indicator function and η is the learning rate.
  3. Output $\hat{f}(\mathbf{x}) = f_M(\mathbf{x})$.

In this work, we used the Least Squares Loss function, simplifying the initialization to $f_0(x) = \bar{y}$, and the residuals to $r_i^{(m)} = y_i - f(\mathbf{x}_i)$.
This iterative approach enables GBT to focus on improving the predictions where prior models underperformed, often resulting in enhanced accuracy.

- A Gaussian process $g(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ is a stochastic process fully specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$, consisting of an infinite collection of random variables, where any finite subset has a multivariate Gaussian distribution [95].
In realistic scenarios, it is often assumed the mean $m(\mathbf{x}) = \mathbf{h}(\mathbf{x})^t \boldsymbol{\beta}$, where $h(\cdot)$ are a set of fixed basis functions that transform the original feature vector $\mathbf{x}$ in $\mathbb{R}^d$ into a new feature vector $\mathbf{h}(\mathbf{x})$ in $\mathbb{R}^p$, where parameters $\boldsymbol{\beta}$ must be inferred by data. Moreover, data are often noisy hence we typically have access only to noisy versions of true function values, $y_i = \mathbf{h}(\mathbf{x}_i)^t \boldsymbol{\beta} + f(\mathbf{x}_i) + \varepsilon_i$, $i = 1, \cdots, n$, where $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k(\mathbf{x}, \mathbf{x}'))$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ represents additive independent and identically distributed (i.i.d.) Gaussian noise with covariance $\sigma_n^2 \mathbf{I}$ [95]. This formulation expresses that the data is close to a global linear model with the residuals being modelled by a GP and the model fitting optimizes over the parameters jointly with the hyperparameters of the covariance function.
The kernel $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, \cdots, n$ encodes the similarity between the input points according to some metric, i.e., points characterized by similar features must have similar target values y and thus, training points that are near to a test point should be informative about the prediction at that point. This is translated in terms of correlation. Indicating with $r = d(\mathbf{x}_i, \mathbf{x}_j)$ any distance between points j, j, the most common kernel function to account for such a similarity between points is the squared exponential (SE)

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{r}{2l^2}\right) \tag{9}$$

where l, also called the length scale parameter, indicates how quickly the correlation between two points drops as their distance increases. This kernel is infinitely differentiable, which means that the GP has mean square derivatives of all orders, and is thus very smooth and such strong smoothness assumptions can be unrealistic for modelling many physical processes. So an alternative option is the Matérn class defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{l}\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu} r}{l}\right) \tag{10}$$

with $\nu > 0, l > 0$ and K_ν is a modified Bessel Function [98] (see section 9.6). Note that for $\nu \to \infty$ we obtain the SE covariance function. For the Matérn class the process $f(\mathbf{x})$ is k-times Mean Square differentiable if and only if $\nu > k$. For the definition of Mean Square Differentiability see Section 4.1.1 in [95]. When $\nu = p + 1/2$ with p non-negative integer, the kernel becomes a product of an exponential and a polynomial of order p, and the most interesting case for machine learning are $p = 3/2$ and $p = 5/2$, for which

$$\begin{aligned} k_{3/2} &= \left(1 + \frac{\sqrt{3} r}{l}\right) \exp\left(-\frac{\sqrt{3} r}{l}\right) \\ k_{5/2} &= \left(1 + \frac{\sqrt{5} r}{l} + \frac{\sqrt{5} r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5} r}{l}\right). \end{aligned}$$

Using standard algebraic manipulation it is possible to evaluate the predictive distribution in the point y_{new} given the new features vector $\mathbf{x}_{new}$

$$P(y_{new}|\mathbf{y}, \mathbf{X}, \mathbf{x}_{new}) = N\Big(h(\mathbf{x}_{new})^t\boldsymbol{\beta} + \mu, \sigma^2 + \Sigma\Big), \tag{11}$$

where

$$\mu = K(\mathbf{x}_{new}^t, \mathbf{X})\Big(K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}\Big)^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\beta}) \tag{12}$$

$$\Sigma = k(\mathbf{x}_{new}, \mathbf{x}_{new}) - K(\mathbf{x}_{new}^t, \mathbf{X})\Big(K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}\Big)^{-1}K(\mathbf{X}, \mathbf{x}_{new}^t) \tag{13}$$

with

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^{n\times 1}, \mathbf{f} = \begin{pmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_n) \end{pmatrix} \in \mathbb{R}^{n\times 1},$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix} \in \mathbb{R}^{n\times d}, \mathbf{H} = \begin{pmatrix} h(\mathbf{x}_1)^t \\ h(\mathbf{x}_2)^t \\ \vdots \\ h(\mathbf{x}_n)^t \end{pmatrix}, \in \mathbb{R}^{n\times p}.$$

$$K(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & k(\mathbf{x}_n, \mathbf{x}_1) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

- Artificial neural networks (ANN), or simply Neural Networks (NN) take inspiration from how the human brain computes complex processes and, specifically, how neurons in our brains work. Similarly to our brain, an NN consists of many units called artificial neurons connected by edges (synapses in our analogy). Each neuron receives a signal from other neurons, processes it and then sends it to other connected neurons. The signal refers to a real number, and each neuron's output is determined by a nonlinear function applied to the sum of its inputs, known as the activation function. The strength of the signal at each connection is controlled by a weight, which is modified during the learning process. A basic feedforward neural network consists of many layers of neurons, and the output of the previous layer serves as the input to neurons in the next layer. Given a nonlinear activation function $f(\cdot)$ (e.g., ReLU, sigmoid, or tanh) a NN can be written as follows:

$$\hat{\mathbf{y}} = f(\mathbf{W}_L \cdot f(\mathbf{W}_{L-1} \cdot \ldots f(\mathbf{W}_1 \cdot \mathbf{X} + \mathbf{b}_1) + \mathbf{b}_{L-1}) + \mathbf{b}_L) \tag{14}$$

where $\mathbf{X}$ is the input matrix, $\mathbf{W}_L$ is the weight matrix associated to layer L, $\mathbf{b}_L$ the L layer bias. Train an NN means finding the weights $\mathbf{W}$ and biases $\mathbf{b}$ that minimize a loss function (e.g., Mean Squared Error for regression), gradient-based methods such as backpropagation are used. Roughly speaking, gradients of the loss function with respect to the weights are calculated and the weights are updated iteratively with the following rule:

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \eta \cdot \nabla L(\mathbf{W}) \tag{15}$$

where η is the learning rate and $\nabla L(\mathbf{W})$ is the gradient of the loss function L with respect to the weights W.

The predictions from the methods described above are used as input for the Stacked Regression. Stacked Regression, first introduced by Breiman [96], is an ensemble learning technique to combine multiple regression models via a meta-regressor. Stacking combines the strengths of individual estimators by using their outputs (i.e., predictions) as inputs for a final estimator. For example, in [99], Stacked Regression outperformed three ML models in downscaling ERA5 air temperature.

Care must be taken when implementing stacked regression, as an improper setup can lead to overfitting. As Breiman highlights, if the base estimators are trained on the entire dataset, the final estimator is likely to overfit the data. A common solution is to use cross-validated predictions: the dataset is divided into k folds, each base estimator is trained on the $k-1$ folds, while the remaining fold is used to compute the predictions. In each iteration, the base estimators generate predictions for the fold that was not used during training, which are then collected across all iterations, stacked together, and then used as input for the final estimator.

Mathematically, the original Stacked Regression proposed by Breiman consists of finding $\hat{\beta}_h$ that minimizes:

$$\sum_{n=1}^{N} (y_n - \sum_{h=1}^{H} \hat{\beta}_h \hat{f}_h(x))^2 \tag{16}$$

where H is the number of base estimators, $\hat{\beta}_h$ are the weights assigned to the predictions of the h-th base estimator, and $\hat{f}_h(x)$ is the prediction of the h-th base estimator for input X, i.e., the stacked estimator is the least squares. This has been generalized to include different types of estimators. Algorithm 1 shows the pseudo-code used in this work to perform Stacked Regression.

Here, the base estimators H are RF, GBT, GPR, and NN , while the final estimator G is the eXtreme Gradient Boosting (XGBoost) [100], an efficient and scalable implementation of the gradient boosting framework that has gained substantial popularity in recent years, largely due to its role in helping teams win nearly every Kaggle structured data competition.

Algorithm 1 Stacking with Cross-Validation

Input:
  Dataset $D = \{(Y_i, X_i)\}_{i=1}^{N}$ where X_n is an input vector
  Number of folds V for cross-validation
  Number of base estimators H
Step 1: Split the dataset into folds:
  Split D into V equal-sized folds $D_1, D_2, \ldots, D_V$
Step 2: Initialize prediction storage:
  Initialize an empty list to store predictions, $\hat{S} = \{\hat{Y}_i^{\,1}, \hat{Y}_i^{\,2}, ..., \hat{Y}_i^{\,H}\}_{i=1}^{N}$.
Step 3: Train base estimators for each fold:
for each fold $v = 1, 2, ..., V$ **do**
    Define training set $D^{(-v)} = D - D_v$ and validation set D_v
  for each model $h = 1, 2, ..., H$ **do**,
      Train the base estimator (level 0) $M_h^{(-v)}$ on $D^{(-v)}$
      Predict $\hat{Y}_v^{(h)}$ using model $M_h^{(-v)}$
      Store prediction $\hat{Y}_v^{(h)}$ in $\hat{S}(v, h)$
  end for
end for
Step 4: Create level 1 learning sample:
  Create $D_{\text{CV}} = \{(Y, \hat{S})\}$
Step 5: Train the final estimator (level 1 learner):
  Train the final estimator G (level 1 learner) on D_{CV}
Step 6: Final prediction:
for each new input X_i in test data X_{test} **do**
  for each model $h = 1, 2, ..., K$ **do**
      Predict $\hat{Y}_i^{(k)}$ using base model M_h
  end for
    Stack predictions into $X_{\text{stacked\_test}} = (\hat{Y}_i^{(1)}, \hat{Y}_i^{(2)}, \ldots, \hat{Y}_i^{(H)})$
    Predict final output using the final estimator $\hat{Y}_i = G(X_{\text{stacked\_test}})$:
end for
Output: Final predictions $\hat{Y}_{\text{test}}$ for test data

2.6. Training

Then, the training dataset built in the Section 2.4 is used as input for the ML algorithms. By conducting an ablation study, we were able to isolate the most impactful covariates for the model. The ablation study involved iteratively removing subsets of covariates and assessing the impact on the RMSE. Variable importance plots, which quantify how much each feature influences the model's accuracy, are also used. The best combination of covariates we found is as follows:

$$LST_{IASI} = f(lat, lon, DEM, DOY, minutes, year, LST_{SEVIRI}, LST_{AVHRR}, CLC, LAI, FVC, NDVI) \quad (17)$$

where DOY (Day of the Year) is acquisition day (from 1 to 365) and minutes refers to the acquisition time expressed in minutes (from 1 to 1440). Since our goal is to predict a monthly product, we removed the seasonal variability by training a model for each month for a total of 12 models for each ML method. LAI and FVC were removed during the winter months (from November to April) since they were not useful in those months. The workflow is as follows:

1. The whole training dataset is used to perform a hyperparameter optimization of the base estimators (RF, GB, GPR, NN). These models will be used again for prediction, see Section 2.7;

2. The training dataset is divided into 5 categories, the base estimators are trained on the first 4/5 of the data using the hyperparameter found in step 1. The last fold is used for prediction. At each iteration, the predictions are concatenated in a vector;
3. Predictions from step 2 are used as input for the final estimator (XGBoost). Hyperparameter optimization for XGBoost is also performed.
4. The trained XGBoost is used to predict the L2 observations and compute the residuals (i.e., the difference between the real IASI L2 observations and the predicted ones) required for residual correction.

More specifically, all the ML models are trained within Matlab using "fitrensemble" for RF and GBT, "fitrgp" for GPR, "fitrnet" for NN. XGBoost is not natively implemented in Matlab (version R2024b), but in [101] there is an implementation of the C++ version. This code works only for classification, but we adapted it for regression implementing also the Matlab Bayesian Hyperparameter optimization algorithm, a faster approach compared to the classic grid search or random search methods. The same hyperparameter optimization algorithm is used for the other ML models. A five-fold cross-validation is used in the hyperparameter optimization step. Moreover, for RF and GBT we optimized the number of trees, maximized the number of decision splits (or branch nodes), minimized the number of leaf node observations, and the learning rate (only for GBT); for the XGBoost learning rate, a maximum depth of the trees, and minimum weight was required in order to create a new node in the tree. For the NN, we followed a more rigorous approach using the formula provided in [102] for a two-hidden-layer NN with m output neurons and N number of samples. The sufficient number of neurons in the first layer is $\sqrt{(m+2)N} + \sqrt{N/(m+2)}$, and in the second it is $m\sqrt{N/(m+2)}$. In our case, we used 300 neurons for the first layer and 60 for the second layer. Finally, for the GPR we used the Matern 3/2 kernel as kernel function.

2.7. Prediction

The aim of our study is to predict the LST on a 0.05° L3 grid. This prediction is carried out using the models trained in Section 2.6 and an L3 dataset (referred to as the test dataset) that has the same spatial resolution. To build the test dataset, we adopted an approach similar to the one outlined in Sections 2.4 and 2.6. For spatial colocalization, instead of using the IASI footprints, we used the grid cells of a 0.05° regular grid placed on our target region, resulting in a total of 4961 pixels. Therefore, we averaged all the covariate values within each 0.05° grid cell. Regarding temporal colocalization, we took advantage of the high temporal resolution of the covariates creating a daily test dataset. Thus, we applied a temporal colocalization similar to that shown in Section 2.4, where instead of using different acquisition days and minutes for each grid cell, we used the following: (1) for a given day calculated the minimum, the maximum, and the average for the acquisition time related to the corresponding IASI observations; (2) used the mean acquisition time as a reference for all covariates except for SEVIRI, where we applied the minimum and maximum acquisition times, averaging all observations within this time window; (3) assigned the current day to the DOY variable and the mean acquisition time to the minutes variable. Pixels with missing values are removed. Then, we used the base estimator trained on the whole training dataset (Step 1 in Section 2.6) to compute daily predictions over the L3 grid. From these intermediate predictions, we predicted the final estimates using the XGBoost model. The final daily predictions are averaged over the whole month. Caution is necessary when using ML algorithms for interpolation. The resulting LST maps contain values where the covariates are available. In the best-case scenario, where covariates have no missing data, the resulting LST maps are also complete. Unfortunately, certain regions, such as the Alps, lack IASI L2 observations, even when considering nine years of data. This means that the

trained models have not seen any observations in those impervious areas, resulting in a prediction based only on the remaining data (which are different). Taking this example to the extreme, it is like training a model on data from the African Savannah and then attempting to use it to predict conditions in the Russian boreal forest. For this reason, we discarded the L3 LST predictions in areas with a total absence of L2 observations.

The final step is a residual correction. As described in [103], residual correction is a post-processing technique that adjusts the final prediction based on the residuals, which represent the error on the training set. Therefore, we built an L3 0.05° residual map by colocalizing the residuals to the L3 grid taking into account the corresponding IASI footprint (which can vary depending on the IASI FOV angle).

Since L2 observations are scattered, we don't have residuals all over the region of interest. For this reason, we used thin-plate smoothing spline (Matlab "tpaps" function) to interpolate the missing values and have a complete residual map. Figure 2 illustrates the complete workflow from the IASI LST retrieval to the final downscaled L3 LST. Panel (a) represents the IASI LST retrieval described in Section 2.3, panel (b) depicts the training phase highlighted in Section 2.6, and panel (c) illustrates the procedure for predicting the final L3 LST, as discussed in Section 2.7.

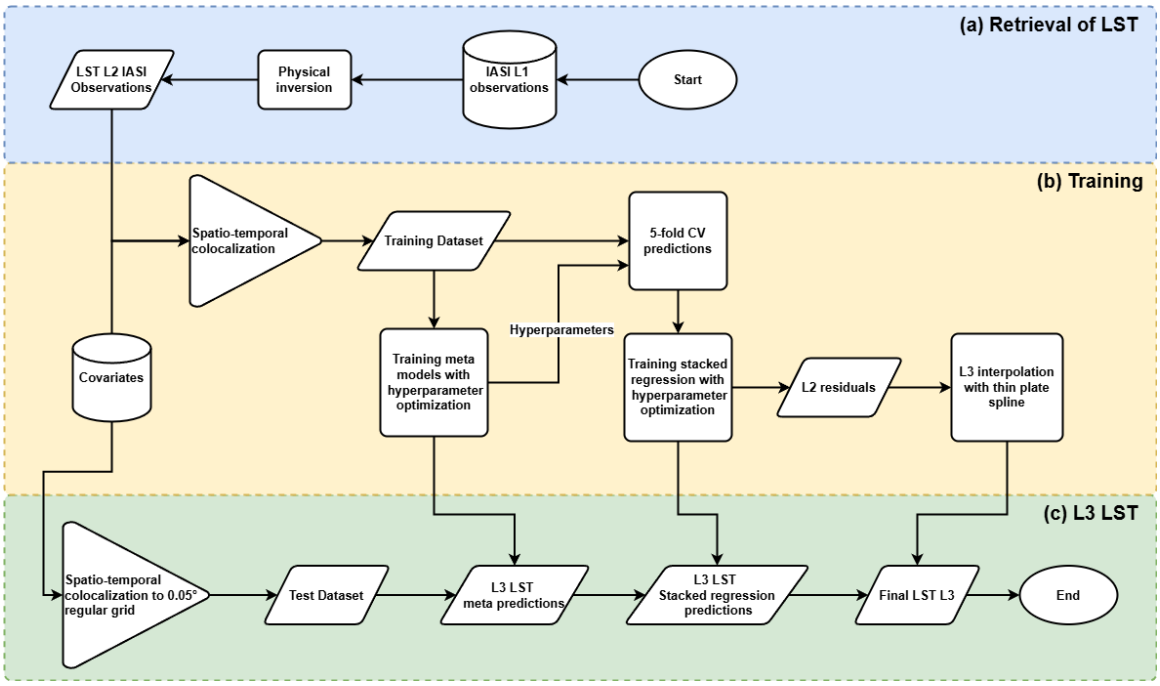

Figure 2. Flowchart of the proposed framework. (**a**) Retrieval of LST; (**b**) Training; (**c**) L3 LST.

3. Results

We applied the methodology described in the previous sections to retrieve the LST over the target region for the year 2022. Note that the drought that hit northern Italy in 2022 was unprecedented in more than two centuries, and is part of a long-term trend of more frequent and severe drought in the area [104].

Figure 3 shows the training observations for the 9 year period 2015–2023 along with the spatial domain used for prediction at 0.05° for the month of August as an example, we again stress that in areas where L2 observations are missing across all years, no predictions were made. Therefore, the mask was created by verifying if at least one L2 observation over all years falls within a 0.05° grid cell.

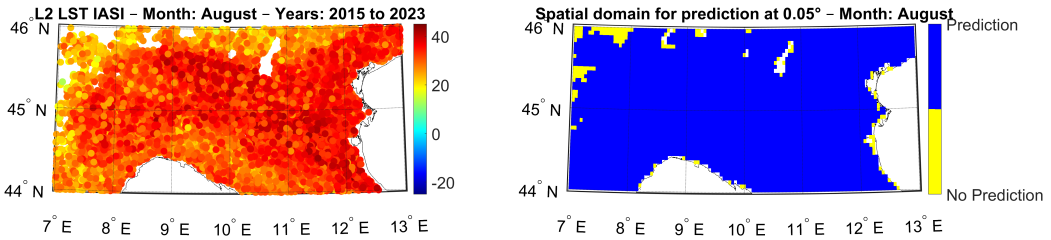

Figure 3. Comparison of L2 IASI observations and the derived prediction mask for August 2022. The left panel shows the spatial distribution of L2 observations across the 9 years of data, while the right panel shows the spatial domain used for prediction.

We first selected the best-performing algorithm among all the trained models. For each month, we computed the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) on our cross-validated predictions of Step 2 in Section 2.6. Since the data used for each prediction in each fold differs from the training set, MAE and RMSE reflect errors evaluated on a test set rather than a training set, thereby minimizing the risk of choosing a model that overfits. Figures 4 and 5 show that the Gradient Boosting is the best among the ML algorithms with MAE of 0.52 and RMSE of 0.71. However, combining these algorithms with Stacked Regression further improved the results, reducing the MAE and RMSE to 0.49 and 0.68, respectively. Hence, Stacked Regression was selected as the candidate model for the L3 predictions, as the errors on the L2 observations are well below 1 K.

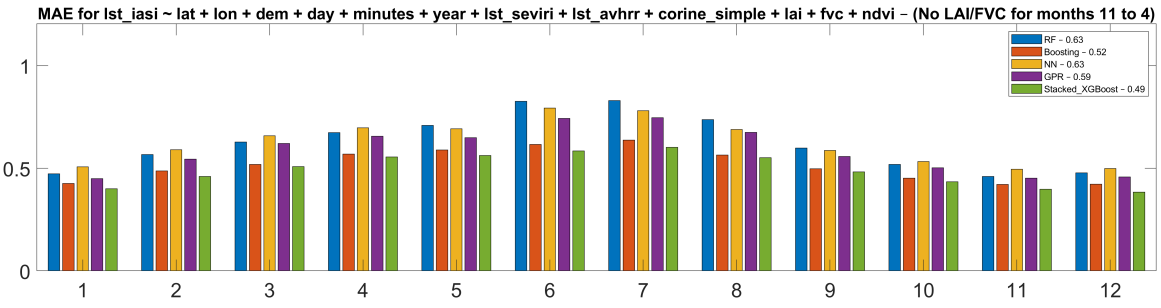

Figure 4. Comparison of MAE cross validated errors of the tested ML algorithms: Random Forest (blue), Boosting (orange), Neural Network (yellow), Gaussian Process Regression (purple), and Stacked Regression (green). The numbers in the legend represent the average MAE for all methods calculated across all months.

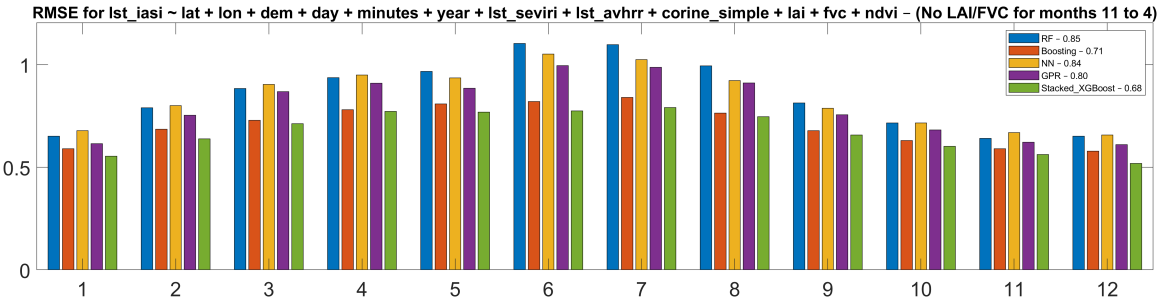

Figure 5. Comparison of RMSE cross validated errors of the tested ML algorithms: Random Forest (blue), Boosting (orange), Neural Network (yellow), Gaussian Process Regression (purple), and Stacked Regression (green). The numbers in the legend represent the average RMSE for all methods calculated across all months.

Following Section 2.7, we generated the L3 LST maps at a resolution of 0.05°, see Figures 6 and 7.

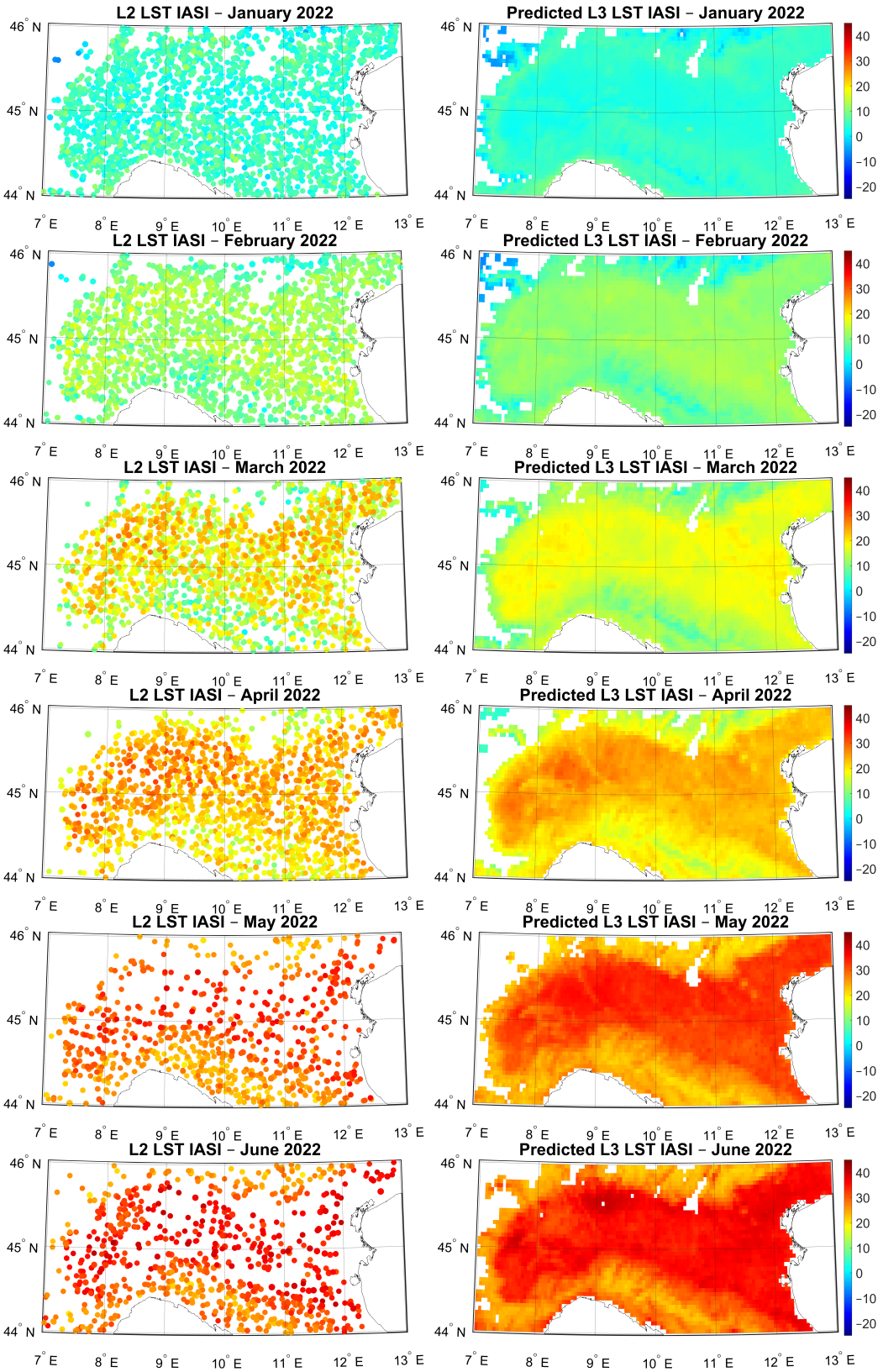

Figure 6. Example of the L3 LST for the months January–June. The first column represents the IASI LST L2 observations, while the second column shows the LST L3 predicted with Stacked Regression; each row represents a month.

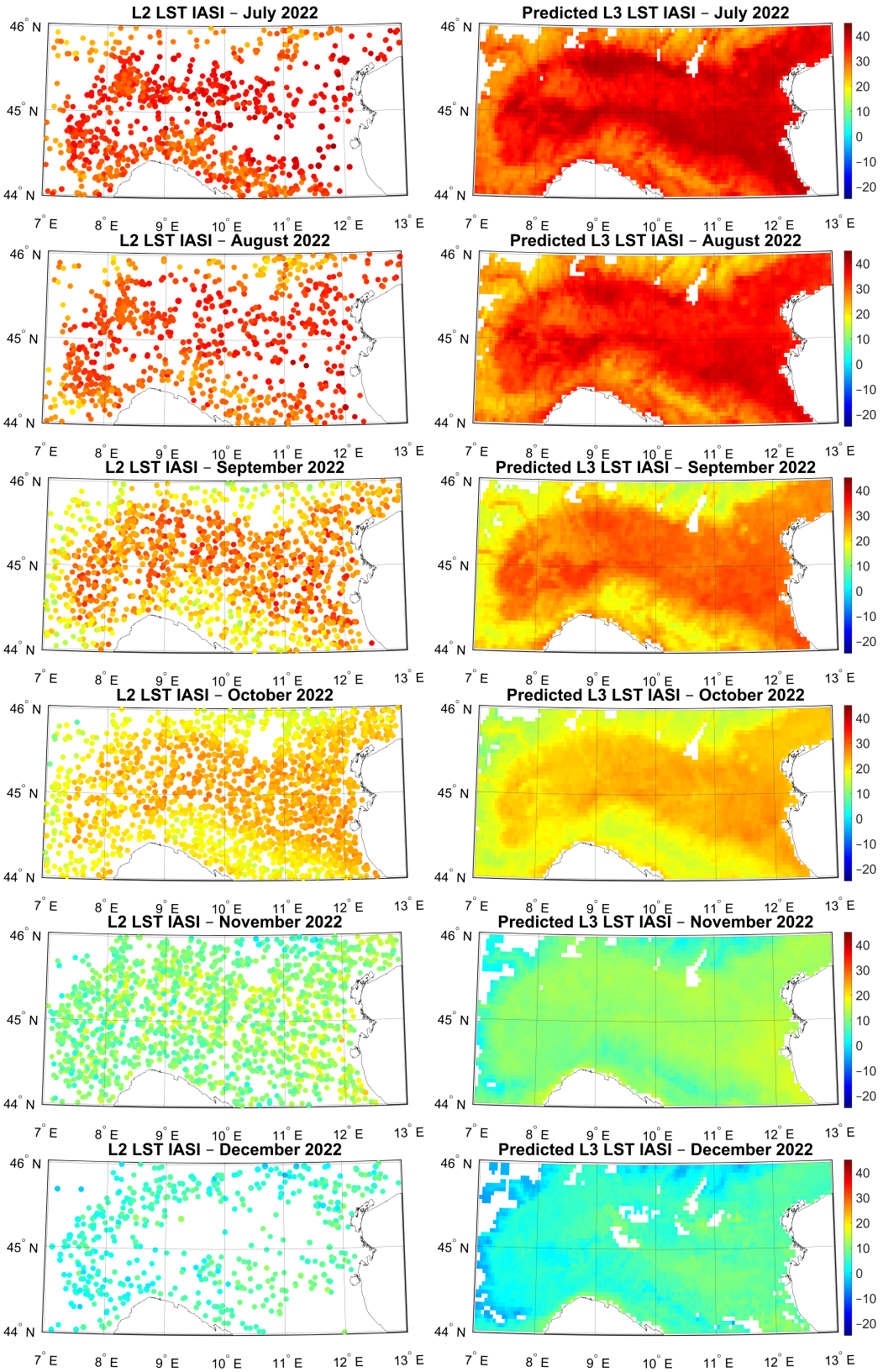

Figure 7. Example of the L3 LST for the months July–December. The first column represents the IASI LST L2 observations, while the second column shows the LST L3 predicted with Stacked Regression; each row represents a month.

There is a strong agreement between the L2 observations and the predicted L3 maps, although the L2 maps are an aggregated product of all L2 observations for a given month (with acquisition times spanning up to 2 h between the first and last observation of a day

over our target area). This explains the higher variability in the L2 observations compared to the L3 maps, which, in contrast, represent a monthly product. This is another clear advantage of this methodology: using a classical interpolation method, such as kriging, on scattered data that varies both spatially and temporally can produce inaccurate maps. For example, if there are only a few observations in a given area, and these observations are acquired two hours after the surrounding pixels, the interpolated LST values for that area will likely be higher than those of the surrounding pixels, resulting in maps that appear "bumpy" or "bubble-shaped". A graphical proof of the accuracy of the predicted values is the high level of detail in the LST maps: based solely on the LST values we can clearly distinguish the plains, the Apennines, the Alps, and even specific features such as valleys and woodlands. For example, in the area around 8°E and 45°N in April, May, September, and October, the LST has lower values compared to the surrounding areas. Upon examining the CLC map, we see that this area, shaped like a horse, corresponds to the Monferrato Hills, a wooded region with hills between 200 and 500 m high. The lower LST values in this area are explained by the cooling effect provided by healthy vegetation.

Note that our proposal retrieves LST in months with low IASI observation density due to the high spatial covariates used for the training phase. See, for example, the months of May, June, July, August, and particularly December. Still, we stress again that in these areas, the retrieval is not completely independent from IASI observations since, in the training phase, we consider only spatial areas with IASI coverage. For example, we focused on August 2022 where we calculated the difference between the prediction and the mean of nine years of AVHRR and SEVIRI LST for the same month. The accuracy of these predictions is illustrated in Figure 8 together with the spatial locations of observations. It is interesting to note that IASI L3 LST retrieval better agrees with SEVIRI LST mean than AVHRR LST mean, perhaps due to a greater spatial homogeneity between IASI and SEVIRI field of views than AVHRR. This behavior is confirmed for the other months too. For the sake of brevity, we only show the plot for March 2022, see Figure 9, the worst month in terms of precipitation and dryness conditions.

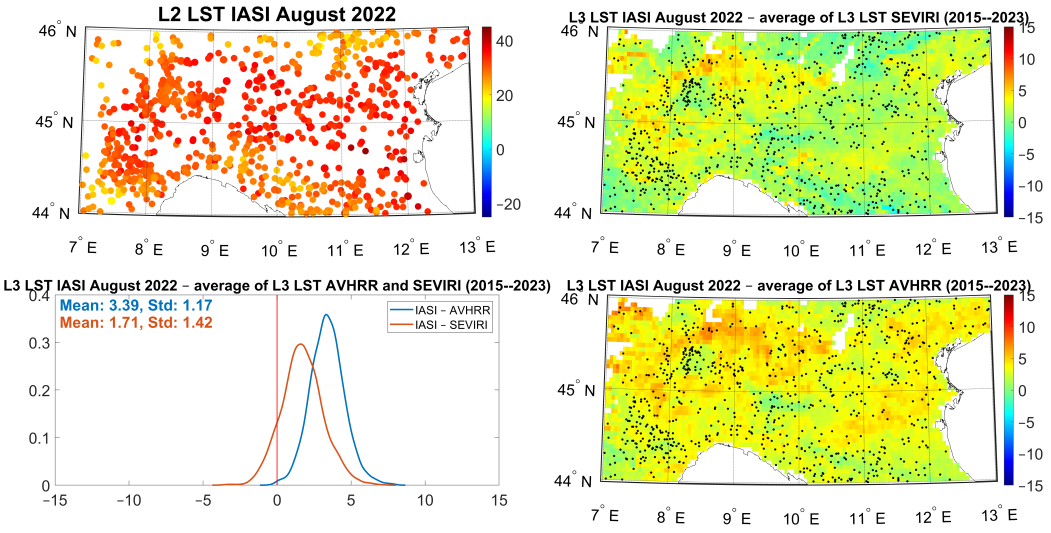

Figure 8. Comparison between the predicted LST for August 2022 and the mean of nine years of AVHRR and SEVIRI data for the same month. The top-left panel shows the IASI L2 observations, while the right panel displays the difference maps with SEVIRI (**top**) and AVHRR (**bottom**) including also the L2 observations represented by the small black dots. The bottom-left panel presents the KDE plot of these differences, including the mean and standard deviation of the errors.

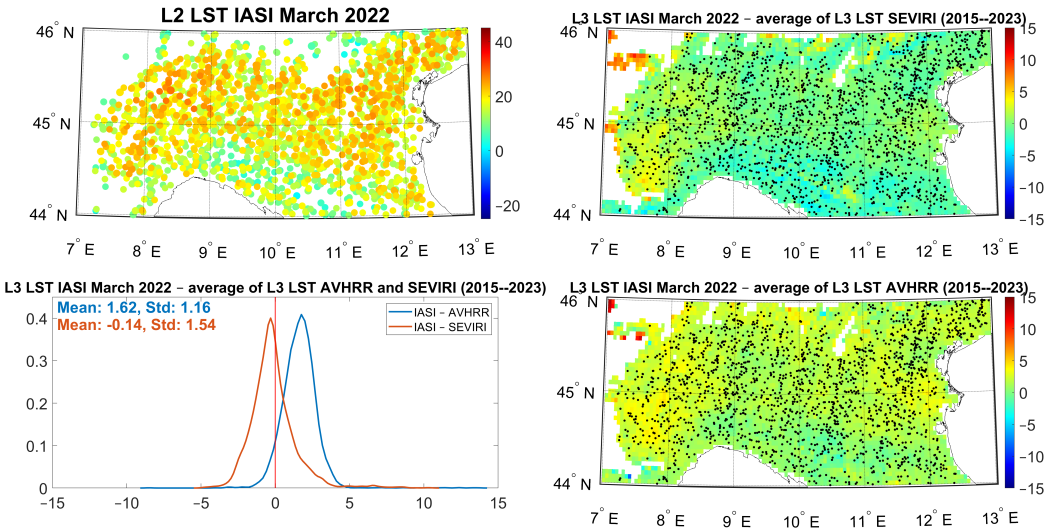

Figure 9. Comparison between the predicted LST for March 2022 and the mean of nine years of AVHRR and SEVIRI data for the same month. The top-left panel shows the IASI L2 observations, while the right panel displays the difference maps with SEVIRI (**top**) and AVHRR (**bottom**) including also the L2 observations represented by the small black dots. The bottom-left panel presents the KDE plot of these differences, including the mean and standard deviation of the errors.

A more robust comparison is given considering the LST derived from MODIS (see Section 2.2.4). Both IASI and MODIS are sensors onboard polar satellites with a close acquisition time (especially for the Terra satellite). The daily observations are first averaged over the month and then interpolated to the 0.05° grid by averaging all MODIS pixels that lie inside each 0.05° grid cell. For each month and year the following was conducted:

- We computed the differences between the two products both at levels L2 and L3. The IASI L2 observations and the corresponding MODIS LST were colocalized using the same approach described in Section 2.4. For each month we also evaluated the bias (average of the errors) and standard deviation. The L2/L3 differences are shown using a Kernel Density Estimate (KDE);
- We fitted a linear model, $LST_{MODIS} \sim LST_{IASI}$, to assess the linear relationship. For each month we also evaluated R^2 index. See Figures 10 and 11.

Although the two products are derived from different sensors, with different acquisition times, and algorithms, our predicted LST exhibits a strong correlation with an R^2 greater than 0.90 for almost all months. Additionally, the standard deviation shows good results, ranging from 1 K to 1.5 K. On the other hand, bias is highly seasonal-dependent with a positive bias during the summer months (indicating higher IASI LST) and a negative bias in the winter months (indicating higher MODIS LST). However, the bias is not only still present in the L2 data but is even more pronounced.

The seasonal variability is also addressed by comparing the colocalized (L2) LSTs from IASI, AVHRR, SEVIRI, and MODIS. Figures 12 and 13 present the KDE plots and boxplots of the differences between the L2 LSTs.

Among the different LSTs, the one derived from AVHRR shows the highest agreement with IASI, exhibiting the highest R^2 values (with the only exception of January and December), the lowest absolute bias (<1 °C) and standard deviation (<2 °C). In contrast, MODIS and SEVIRI have biases that vary significantly throughout the year, with a very negative bias during winter months and a positive bias in the summer months, consistent with patterns observed for the L3 product in Figures 10 and 11. Additionally, their standard deviation values are higher than those of AVHRR but do not show any seasonal patterns. SEVIRI presents higher R^2 values compared to MODIS, similar to those of AVHRR.

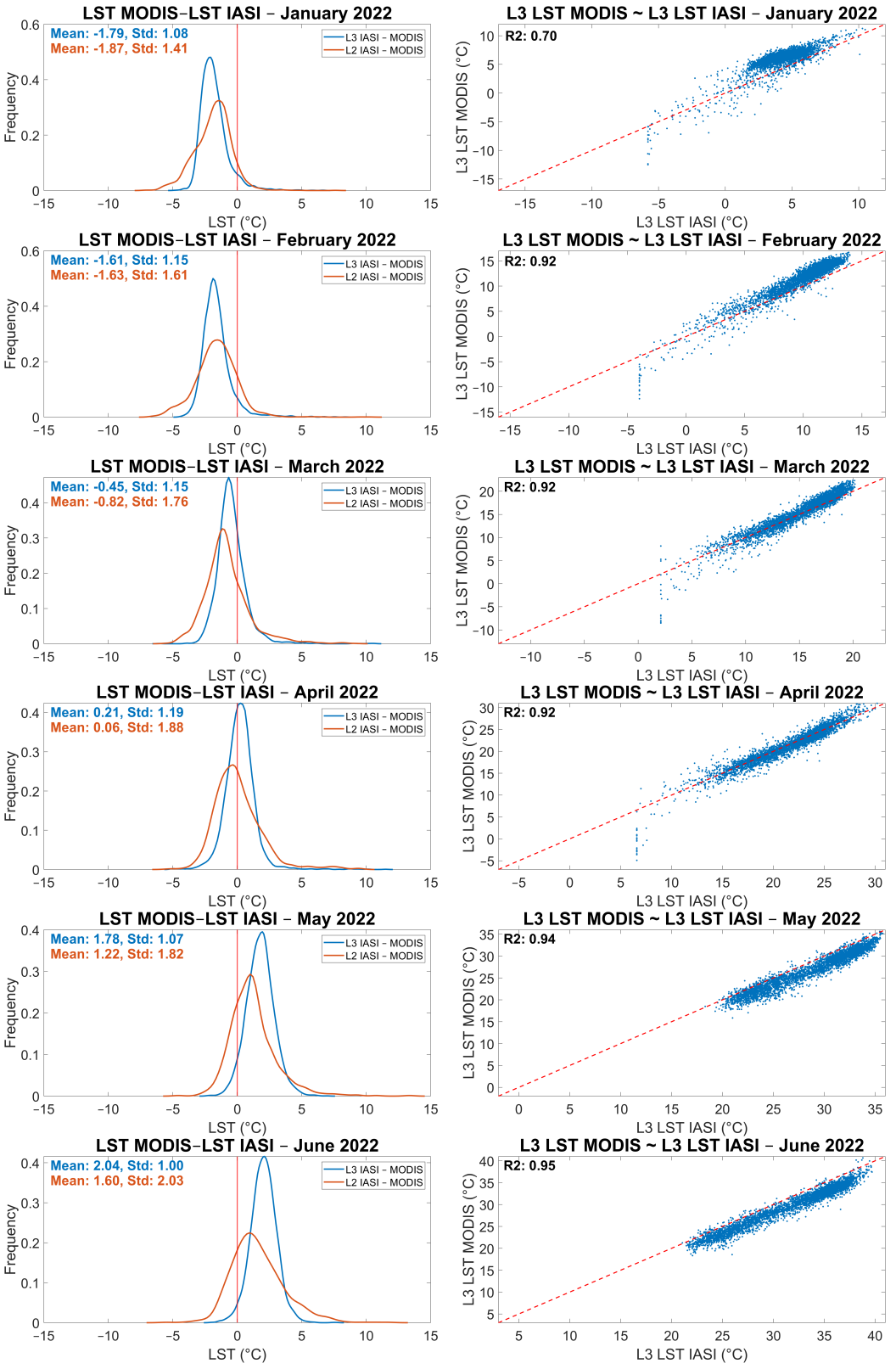

Figure 10. L2/L3 Differences IASI - MODIS for the months January–June. The first column displays the L2/L3 differences using KDE plots with the mean and standard deviation: the red curves display the L2 errors, and the blue curves display the L3 errors. The second column shows the scatterplots between the predicted IASI L3 LST values and the MODIS L3 LST values, the linear fits, and the R^2 indexes.

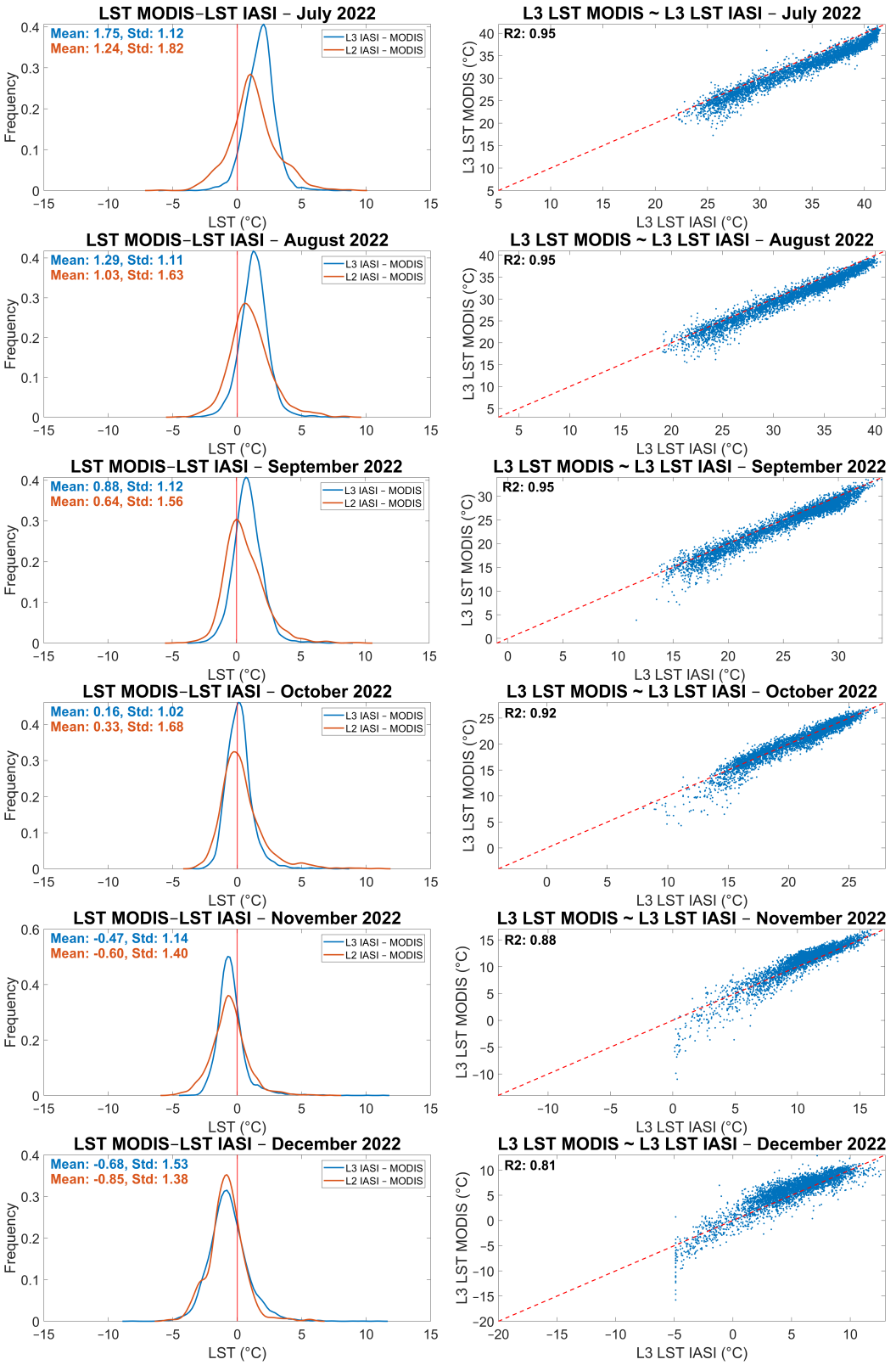

Figure 11. L2/L3 Differences IASI - MODIS for the months July–December. The first column displays the L2/L3 differences using KDE plots with the mean and standard deviation: the red curves display the L2 errors, and the blue curves display the L3 errors. The second column shows the scatterplots between the predicted IASI L3 LST values and the MODIS L3 LST values, the linear fits, and the R^2 indexes.

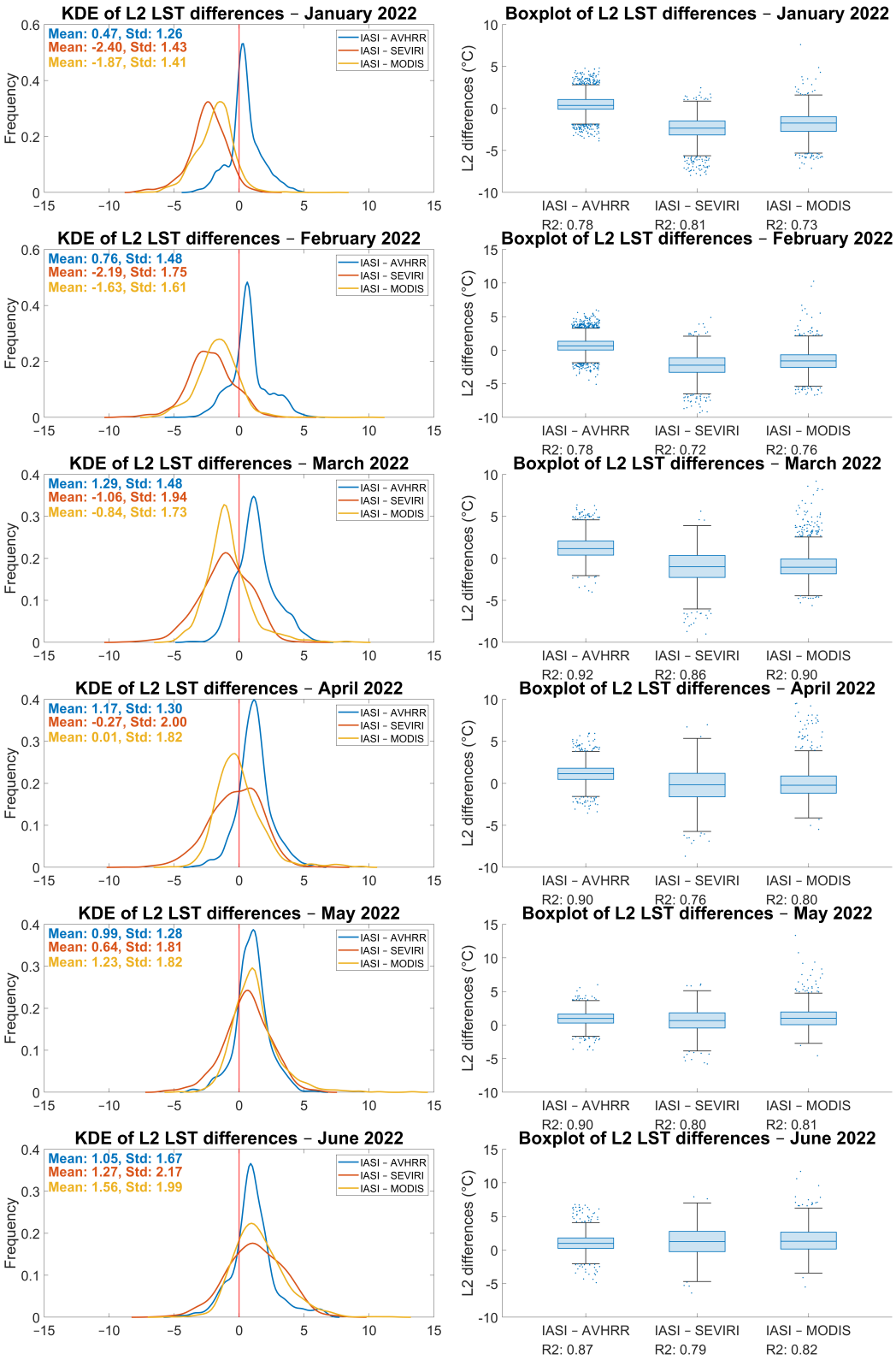

Figure 12. Comparison LSTs from IASI, AVHRR, SEVIRI and MODIS for the months January–June. The first column displays the L2 differences using KDE plots with the mean and standard deviation: the blue curves represent the L2 differences between IASI and AVHRR, the red curves represent the L2 differences between IASI and SEVIRI, and the yellow curves represent the L2 differences between IASI and MODIS. The second column shows the same differences using boxplots, with the R^2 indexes included on the *x*-axis.

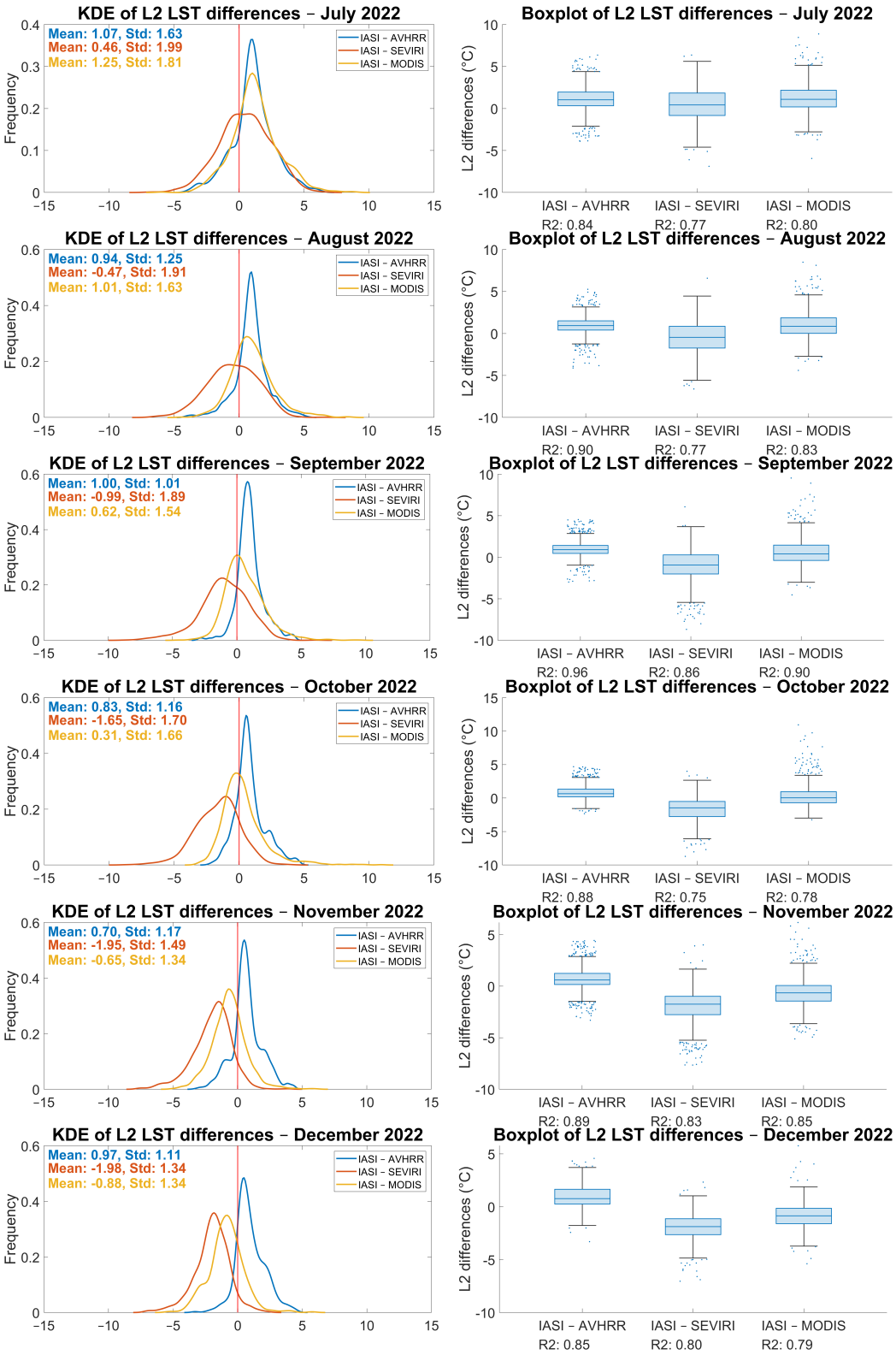

Figure 13. Comparison LSTs from IASI, AVHRR, SEVIRI and MODIS for the months July–December. The first column displays the L2 differences using KDE plots with the mean and standard deviation: the blue curves represent the L2 differences between IASI and AVHRR, the red curves represent the L2 differences between IASI and SEVIRI, and the yellow curves represent the L2 differences between IASI and MODIS. The second column shows the same differences using boxplots, with the R^2 indexes included on the *x*-axis.

4. Discussion

The innovation brought by this work is two-fold. The first is the LST retrieval from the IASI sensor using the physical retrieval package ϕ-IASI. This approach simultaneously retrieves LST, LSE, and atmospheric composition starting from a priori information on surface and atmosphere. The second innovation involves applying an ML downscaling–fusion methodology to both interpolate missing LST values and enhance spatial resolution, achieving promising results. While the literature contains hundreds of studies on downscaling LST, relatively little attention has been given to LST from IASI, whose naturally irregular and scattered observations add another layer of complexity.

We first tested RF, GBT, NN, and GPR, finding that GBT exhibits the lowest errors, with an average cross-validated MAE and RMSE of 0.52 and 0.71, respectively. A classical approach would be to use GBT as the final predictor, since it showed the lowest errors. However, a more accurate predictor can be obtained by combining (stacking) these predictors using Stacked Regression. The results presented in [96] demonstrated that Stacking never performs worse than selecting the single best predictor, and that the biggest improvement comes when dissimilar sets of predictors were stacked. For example, stacking k different NNs is unlikely to yield significant improvement since the predictors are learning from the data in similar ways. In our case, we applied Stacked Regression on predictors that each operates in its own unique way and, indeed, results further improved, lowering the cross-validated MAE and RMSE to 0.49 and 0.68, respectively.

The generated L3 maps are then compared against the LST from MODIS (specifically the MOD11A1 product), showing a standard deviation lower than 1.5 K and a bias that fluctuates significantly across different months. A portion of these standard deviation values are attributed to the Alps, a region where IASI observations are sparse or entirely absent due to the constant presence of snow. Consequently, the model lacks sufficient training in this area. Instead, the seasonal bias may be explained by the MODIS split-window retrieval algorithm, which uses a fixed emissivity, in contrast to our methodology, which simultaneously retrieves both LST and emissivity. We also assessed the goodness of the downscaling method by comparing the IASI L2 LST with the colocalized MODIS LST. The same seasonal bias and a slightly higher standard deviation persist, indicating that the differences arise from the products themselves rather than our downscaling method, which, by contrast, slightly reduced the differences. The same seasonal trend was observed in the comparison with the L2 SEVIRI, while AVHRR demonstrated the highest agreement with IASI, as both are on board the same platform. However, due to the strong spatial and temporal variability of LST, comparing IASI and MODIS data with a time difference of one or more hours can result in temperature differences of several degrees or more in certain regions. It is also important to note that in order to maintain a sufficient number of data points for comparison, we included MODIS, AVHRR, and SEVIRI pixels with LST uncertainties up to 2 K (nominal quality), as stated in Section 2.2.4. Furthermore, this time difference not only influences LST values due to the diurnal LST cycle but also by the amount of pixels averaged within the IASI footprint, which depends on cloud cover. For example, consider an IASI observation taken at a specific time under clear-sky conditions. Since the LSTs from the other sensors (e.g., MODIS) have a much finer resolution, multiple observations fall within the IASI footprint. If the sky remains completely cloud-free, we will average always the same number of observations. However, during the time window between the IASI observation and the corresponding LST observations from other sensors, clouds could move and partially cover the footprint. In effect, cloud coverage and the same land coverage can affect the comparison among different sensors because the other platforms overpass the given area at various times, and cloud dynamics is much faster than the delay among the platforms themselves. The situation explains why our results better

agree with the AVHRR, a companion instrument of IASI onboard the identical MetOp satellites. This would reduce the number of clear-sky observations used in the averaging process, resulting in LST values that represent only a portion of the footprint. This could explain the seasonal bias observed with SEVIRI and MODIS; during winter months, clouds are much more prevalent, reducing the number of LST observations averaged for each IASI footprint. To substantiate this claim, AVHRR, which shares the same platform and observes the exact same scene, does not show this seasonal bias. Despite that, we observe a strong correlation between the two L3 products, with an R^2 value reaching 0.95 in some months. The L2 comparisons among the four LSTs also revealed a strong correlation, with AVHRR showing the highest R^2, followed by SEVIRI and MODIS. However, the R^2 values are still the highest in the L3 comparison.

Moreover, this procedure is also suitable for gap-filling. For instance, in August 2022 (see Figure 7), predictions were made even in areas lacking L2 observations for the considered year. These predictions are possible because the trained models are taking advantage of L2 observations from other years. While it is expected that predictions in areas with fewer L2 observations are more influenced by the covariates, Figure 8 indicates that some areas show large differences despite the limited number of observations. This suggests that in these regions, the predicted LST is not merely an interpolation of the LST covariates but still incorporates information from IASI observations of other years.

5. Conclusions

The greatest advantage of this work is the generation of a high-resolution LST product from IASI, a sensor capable of retrieving numerous atmospheric parameters, which facilitates the study of the relationships between LST and other atmospheric variables. For instance, LST from IASI is used to compute the Water Deficit Index (WDI), an index proposed by Masiello et al. [89,90], defined as the linear difference between LST and dew point temperature (Tw). This index has already proven useful for detecting drought events [69,89–91], and preliminary efforts have been made to downscale it using machine learning [105]. Still, directly validating the WDI is challenging due to the lack of a ground truth. Most LST products are derived from infrared sensors, while the only Tw product with adequate spatial resolution is the ERA5-Land dataset that is derived from a global atmospheric reanalysis model. Combining LST and Tw from different algorithms or sensors (e.g., Tw from ERA5 and LST from MODIS) will introduce uncertainties. On the other hand, LST and Tw individually are more straightforward to validate. Future work will focus on applying this methodology to the IASI Tw retrieved using the same physical inversion algorithm. This will ultimately aim to create a high-resolution WDI in which both components are validated against other sensors or ground-based observations. Moreover, the strategy we have developed for WDI exploits the capability of IASI to retrieve LST and Tw simultaneously, and as such, the study has been performed in perspective of the new missions, IASI-NG and MTG-IRS, which will continue to contribute to the European project of Earth Observations from polar and geostationary orbit for the future years, but also offer significant enhancement of the current suite of instruments on Metop and MSG satellites, with improved spectral resolution, improved spatial coverage, and repeat time.

Author Contributions: Conceptualization, I.D.F.; methodology, I.D.F. and G.M.; software, F.D.R.; validation, F.D.R.; formal analysis, G.M. and C.S.; investigation, F.D.R. and P.P.; data curation, F.D.R. and P.P.; writing—original draft preparation, F.D.R.; writing—review and editing, I.D.F. and G.M. and C.S.; supervision, I.D.F. and C.S.; funding acquisition, G.M. and C.S. All authors have read and agreed to the published version of the manuscript.

Funding: G.M., P.P. and C.S. acknowledge the project 'Tech4You—Technologies for Climate Change Adaptation and Quality of Life Improvement' (CUP C43C22000400006) funded by the Italian Ministry of University and Research.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABI	Advanced Baseline Imager
AVHRR	Advanced Very High-Resolution Radiometer
CLC	Corine Land Cover
CNN	Convolutional Neural Networks
DEM	Digital Elevation Model
DOY	Day of the Year
EPS	European Polar System
FOR	Fields Of Regard
FVC	Fractional Vegetation Cover
GBT	Gradient-Boosted Trees
GOES	Geostationary Operational Environmental Satellites
GPR	Gaussian Process Regression
IASI	Infrared Atmospheric Sounding Interferometer
IFOV	Instantaneous Fields Of View
IQR	Interquartile Range
KDE	Kernel Density Estimate
LAI	Leaf Area Index
LSE	Land Surface Emissivity
LSA SAF	Satellite Application Facility on Land Surface Analysis
LST	Land Surface Temperature
MAE	Mean Absolute Error
ML	Machine Learning
MIR	Mid-infrared
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
RF	Random Forest
RMSE	Root Mean Squared Error
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SLSTR	Sea and Land Surface Temperature Radiometer
SW	Split-Window
TES	Temperature and Emissivity Separation
TIR	Thermal Infrared
Tw	Dew Point Temperature
VIIRS	Visible Infrared Imaging Radiometer Suite
WDI	Water Deficit Index
XGBoost	eXtreme Gradient Boosting

References

1. Guillevic, P.; Göttsche, F.; Nickeson, J.; Hulley, G.; Ghent, D.; Yu, Y.; Trigo, I.; Hook, S.; Sobrino, J.; Remedios, J.; et al. Land surface temperature product validation best practice protocol. Version 1.1. *Best Pract. Satell.-Deriv. Land Prod. Valid.* **2018**, *60*. [CrossRef]
2. Li, Z.L.; Tang, B.H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [CrossRef]
3. Zhan, W.; Chen, Y.; Zhou, J.; Wang, J.; Liu, W.; Voogt, J.; Zhu, X.; Quan, J.; Li, J. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. *Remote Sens. Environ.* **2013**, *131*, 119–139. [CrossRef]
4. Hook, S.J.; Gabell, A.R.; Green, A.A.; Kealy, P.S. A comparison of techniques for extracting emissivity information from thermal infrared data for geologic studies. *Remote Sens. Environ.* **1992**, *42*, 123–135. [CrossRef]
5. Kealy, P.S.; Hook, S.J. Separating temperature and emissivity in thermal infrared multispectral scanner data: Implications for recovering land surface temperatures. *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 1155–1164. [CrossRef]
6. Dash, P.; Göttsche, F.M.; Olesen, F.S.; Fischer, H. Land surface temperature and emissivity estimation from passive sensor data: Theory and practice-current trends. *Int. J. Remote Sens.* **2002**, *23*, 2563–2594. [CrossRef]
7. Gillespie, A.; Rokugawa, S.; Matsunaga, T.; Cothern, J.; Hook, S.; Kahle, A. A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1113–1126. [CrossRef]
8. Li, Z.L.; Wu, H.; Duan, S.B.; Zhao, W.; Ren, H.; Liu, X.; Leng, P.; Tang, R.; Ye, X.; Zhu, J.; et al. Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications. *Rev. Geophys.* **2023**, *61*, e2022RG000777. [CrossRef]
9. Qin, Z.; Karnieli, A.; Berliner, P. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. *Int. J. Remote Sens.* **2001**, *22*, 3719–3746. [CrossRef]
10. Jiménez-Muñoz, J.C.; Sobrino, J.A. A generalized single-channel method for retrieving land surface temperature from remote sensing data. *J. Geophys. Res. Atmos.* **2003**, *108*, 4688. [CrossRef]
11. Jimenez-Munoz, J.C.; Cristobal, J.; Sobrino, J.A.; Sòria, G.; Ninyerola, M.; Pons, X. Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *47*, 339–349. [CrossRef]
12. McMillin, L.M. Estimation of sea surface temperatures from two infrared window measurements with different absorption. *J. Geophys. Res.* **1975**, *80*, 5113–5117. [CrossRef]
13. Wan, Z.; Dozier, J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 892–905. [CrossRef]
14. Liu, Y.; Yu, Y.; Yu, P.; Wang, H.; Rao, Y. Enterprise LST Algorithm Development and Its Evaluation with NOAA 20 Data. *Remote Sens.* **2019**, *11*, 2003. [CrossRef]
15. Ghent, D.J.; Corlett, G.K.; Göttsche, F.M.; Remedios, J.J. Global Land Surface Temperature From the Along-Track Scanning Radiometers. *J. Geophys. Res. Atmos.* **2017**, *122*, 12167–12193. [CrossRef]
16. Yu, Y.; Tarpley, D.; Privette, J.L.; Goldberg, M.D.; Rama Varma Raja, M.K.; Vinnikov, K.Y.; Xu, H. Developing Algorithm for Operational GOES-R Land Surface Temperature Product. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 936–951. [CrossRef]
17. Trigo, I.F.; Monteiro, I.T.; Olesen, F.; Kabsch, E. An assessment of remotely sensed land surface temperature. *J. Geophys. Res. Atmos.* **2008**, *113*, *D17*. [CrossRef]
18. Chen, Y.; Duan, S.B.; Ren, H.; Labed, J.; Li, Z.L. Algorithm development for land surface temperature retrieval: Application to Chinese Gaofen-5 data. *Remote Sens.* **2017**, *9*, 161. [CrossRef]
19. Ermida, S.L.; Trigo, I.F.; DaCamara, C.C.; Göttsche, F.M.; Olesen, F.S.; Hulley, G. Validation of remotely sensed surface temperature over an oak woodland landscape—The problem of viewing and illumination geometries. *Remote Sens. Environ.* **2014**, *148*, 16–27. [CrossRef]
20. Ye, X.; Ren, H.; Liu, R.; Qin, Q.; Liu, Y.; Dong, J. Land surface temperature estimate from Chinese Gaofen-5 satellite data using split-window algorithm. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5877–5888. [CrossRef]
21. NASA. Land Surface Temperature and Emissivity Product (MxD21) Algorithm Theoretical Basis Document, 2016. Available online: https://modis-land.gsfc.nasa.gov/pdf/MOD21_LST&E_user_guide_C61.pdf (accessed on 28 November 2024).
22. Wan, Z.; Li, Z.L. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 980–996. [CrossRef]
23. Masiello, G.; Serio, C. Simultaneous physical retrieval of surface emissivity spectrum and atmospheric parameters from infrared atmospheric sounder interferometer spectral radiances. *Appl. Opt.* **2013**, *52*, 2428–2446. [CrossRef]
24. Masiello, G.; Serio, C.; Venafra, S.; Liuzzi, G.; Poutier, L.; Göttsche, F.M. Physical Retrieval of Land Surface Emissivity Spectra from Hyper-Spectral Infrared Observations and Validation with In Situ Measurements. *Remote Sens.* **2018**, *10*, 976. [CrossRef]
25. Mastro, P.; Masiello, G.; Serio, C.; Cimini, D.; Ricciardelli, E.; Paola, F.D.; Hultberg, T.; August, T.; Romano, F. Combined IASI-NG and MWS Observations for the Retrieval of Cloud Liquid and Ice Water Path: A Deep Learning Artificial Intelligence Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3313–3322. [CrossRef]

26. Masiello, G.; Serio, C.; De Feis, I.; Amoroso, M.; Venafra, S.; Trigo, I.F.; Watts, P. Kalman filter physical retrieval of surface emissivity and temperature from geostationary infrared radiances. *Atmos. Meas. Tech.* **2013**, *6*, 3613–3634. [CrossRef]
27. Masiello, G.; Serio, C.; Venafra, S.; Liuzzi, G.; Göttsche, F.; F. Trigo, I.; Watts, P. Kalman filter physical retrieval of surface emissivity and temperature from SEVIRI infrared channels: A validation and intercomparison study. *Atmos. Meas. Tech.* **2015**, *8*, 2981–2997. [CrossRef]
28. Masiello, G.; Serio, C.; Venafra, S.; Poutier, L.; Göttsche, F.M. SEVIRI Hyper-Fast Forward Model with Application to Emissivity Retrieval. *Sensors* **2019**, *19*, 1532. [CrossRef] [PubMed]
29. Pu, R.; Bonafoni, S. Thermal infrared remote sensing data downscaling investigations: An overview on current status and perspectives. *Remote Sens. Appl. Soc. Environ.* **2023**, *29*, 100921. [CrossRef]
30. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *6*, e5518. [CrossRef] [PubMed]
31. Sekulić, A.; Kilibarda, M.; Heuvelink, G.B.; Nikolić, M.; Bajat, B. Random Forest Spatial Interpolation. *Remote Sens.* **2020**, *12*, 1687. [CrossRef]
32. Li, J.; Heap, A.D.; Potter, A.; Daniell, J.J. Application of machine learning methods to spatial interpolation of environmental variables. *Environ. Model. Softw.* **2011**, *26*, 1647–1659. [CrossRef]
33. Taripanah, F.; Ranjbar, A. Quantitative analysis of spatial distribution of land surface temperature (LST) in relation Ecohydrological, terrain and socio- economic factors based on Landsat data in Mountainous area. *Adv. Space Res.* **2021**, *68*, 3622–3640. [CrossRef]
34. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141. [CrossRef]
35. Li, W.; Ni, L.; Li, Z.L.; Duan, S.B.; Wu, H. Evaluation of Machine Learning Algorithms in Spatial Downscaling of MODIS Land Surface Temperature. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2299–2307. [CrossRef]
36. Ebrahimy, H.; Azadbakht, M. Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Comput. Geosci.* **2019**, *124*, 93–102. [CrossRef]
37. Yang, Y.; Cao, C.; Pan, X.; Li, X.; Zhu, X. Downscaling Land Surface Temperature in an Arid Area by Using Multiple Remote Sensing Indices with Random Forest Regression. *Remote Sens.* **2017**, *9*, 789. [CrossRef]
38. Yoo, C.; Im, J.; Cho, D.; Lee, Y.; Bae, D.; Sismanidis, P. Downscaling MODIS nighttime land surface temperatures in urban areas using ASTER thermal data through local linear forest. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *110*, 102827. [CrossRef]
39. Li, X.; Zhang, G.; Zhu, S.; Xu, Y. Step-By-Step Downscaling of Land Surface Temperature Considering Urban Spatial Morphological Parameters. *Remote Sens.* **2022**, *14*, 3038. [CrossRef]
40. Lezama Valdes, L.M.; Katurji, M.; Meyer, H. A Machine Learning Based Downscaling Approach to Produce High Spatio-Temporal Resolution Land Surface Temperature of the Antarctic Dry Valleys from MODIS Data. *Remote Sens.* **2021**, *13*, 4673. [CrossRef]
41. Nguyen, B.M.; Tian, G.; Vo, M.T.; Michel, A.; Corpetti, T.; Granero-Belinchon, C. Convolutional Neural Network Modelling for MODIS Land Surface Temperature Super-Resolution. In Proceedings of the 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 29 August–2 September 2022. [CrossRef]
42. Liang, M.; Zhang, L.; Wu, S.; Zhu, Y.; Dai, Z.; Wang, Y.; Qi, J.; Chen, Y.; Du, Z. A High-Resolution Land Surface Temperature Downscaling Method Based on Geographically Weighted Neural Network Regression. *Remote Sens.* **2023**, *15*, 1740. [CrossRef]
43. Choe, Y.J.; Yom, J.H. Downscaling of MODIS land surface temperature to LANDSAT scale using multi-layer perceptron. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2017**, *35*, 313–318. [CrossRef]
44. Guo, S.; Li, M.; Li, Y.; Chen, J.; Zhang, H.K.; Sun, L.; Wang, J.; Wang, R.; Yang, Y. The Improved U-STFM: A Deep Learning-Based Nonlinear Spatial-Temporal Fusion Model for Land Surface Temperature Downscaling. *Remote Sens.* **2024**, *16*, 322. [CrossRef]
45. Ebrahimy, H.; Aghighi, H.; Azadbakht, M.; Amani, M.; Mahdavi, S.; Matkan, A.A. Downscaling MODIS Land Surface Temperature Product Using an Adaptive Random Forest Regression Method and Google Earth Engine for a 19-Years Spatiotemporal Trend Analysis Over Iran. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2103–2112. [CrossRef]
46. Chybicki, A.; Łubniewski, Z. Optimized AVHRR land surface temperature downscaling method for local scale observations: Case study for the coastal area of the Gulf of Gdańsk. *Open Geosci.* **2017**, *9*, 419–435. [CrossRef]
47. Stathopoulou, M.; Cartalis, C. Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sens. Environ.* **2009**, *113*, 2592–2605. [CrossRef]
48. Xu, J.; Zhang, F.; Jiang, H.; Hu, H.; Zhong, K.; Jing, W.; Yang, J.; Jia, B. Downscaling Aster Land Surface Temperature over Urban Areas with Machine Learning-Based Area-To-Point Regression Kriging. *Remote Sens.* **2020**, *12*, 1082. [CrossRef]
49. Hu, Y.; Tang, R.; Jiang, X.; Li, Z.L.; Jiang, Y.; Liu, M.; Gao, C.; Zhou, X. A physical method for downscaling land surface temperatures using surface energy balance theory. *Remote Sens. Environ.* **2023**, *286*, 113421. [CrossRef]
50. Rodriguez-Galiano, V.; Pardo-Iguzquiza, E.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Downscaling Landsat 7 ETM+ thermal imagery using land surface temperature and NDVI images. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 515–527. [CrossRef]

51. Dong, P.; Gao, L.; Zhan, W.; Liu, Z.; Li, J.; Lai, J.; Li, H.; Huang, F.; Tamang, S.K.; Zhao, L. Global comparison of diverse scaling factors and regression models for downscaling Landsat-8 thermal data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 44–56. [CrossRef]
52. Xu, S.; Zhao, Q.; Yin, K.; He, G.; Zhang, Z.; Wang, G.; Wen, M.; Zhang, N. Spatial Downscaling of Land Surface Temperature Based on a Multi-Factor Geographically Weighted Machine Learning Model. *Remote Sens.* **2021**, *13*, 1186. [CrossRef]
53. Abunnasr, Y.; Mhawej, M. Towards a combined Landsat-8 and Sentinel-2 for 10-m land surface temperature products: The Google Earth Engine monthly Ten-ST-GEE system. *Environ. Model. Softw.* **2022**, *155*, 105456. [CrossRef]
54. Onačillová, K.; Gallay, M.; Paluba, D.; Péliová, A.; Tokarčík, O.; Laubertová, D. Combining Landsat 8 and Sentinel-2 Data in Google Earth Engine to Derive Higher Resolution Land Surface Temperature Maps in Urban Environment. *Remote Sens.* **2022**, *14*, 4076. [CrossRef]
55. Su, Q.; Yao, Y.; Chen, C.; Chen, B. Generating a 30 m Hourly Land Surface Temperatures Based on Spatial Fusion Model and Machine Learning Algorithm. *Sensors* **2024**, *24*, 7424. [CrossRef] [PubMed]
56. Zakšek, K.; Oštir, K. Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sens. Environ.* **2012**, *117*, 114–124. [CrossRef]
57. Keramitsoglou, I.; Kiranoudis, C.T.; Weng, Q. Downscaling Geostationary Land Surface Temperature Imagery for Urban Analysis. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1253–1257. [CrossRef]
58. Lee, J.; Berkelhammer, M.; Wilson, M.D.; Love, N.; Cintron, R. Urban Land Surface Temperature Downscaling in Chicago: Addressing Ethnic Inequality and Gentrification. *Remote Sens.* **2024**, *16*, 1639. [CrossRef]
59. Safieddine, S.; Parracho, A.C.; George, M.; Aires, F.; Pellet, V.; Clarisse, L.; Whitburn, S.; Lezeaux, O.; Thépaut, J.N.; Hersbach, H.; et al. Artificial Neural Networks to Retrieve Land and Sea Skin Temperature from IASI. *Remote Sens.* **2020**, *12*, 2777. [CrossRef]
60. Aires, F.; Boucher, E.; Pellet, V. Convolutional neural networks for satellite remote sensing at coarse resolution. Application for the SST retrieval using IASI. *Remote Sens. Environ.* **2021**, *263*, 112553. [CrossRef]
61. Boucher, E.; Aires, F.; Pellet, V. Towards a new generation of artificial-intelligence-based infrared atmospheric sounding interferometer retrievals of surface temperature: Part I—Methodology. *Q. J. R. Meteorol. Soc.* **2023**, *149*, 1180–1196. [CrossRef]
62. Boucher, E.; Aires, F. Towards a new generation of artificial-intelligence-based infrared atmospheric sounding interferometer retrievals of surface temperature: Part II—Assessment. *Q. J. R. Meteorol. Soc.* **2023**, *149*, 1593–1611. [CrossRef]
63. Amato, U.; Masiello, G.; Serio, C.; Viggiano, M. The σ-IASI code for the calculation of infrared atmospheric radiance and its derivatives. *Environ. Model. Softw.* **2002**, *17*, 651–667. [CrossRef]
64. Serio, C.; Masiello, G.; Liuzzi, G.; Cersosimo, A.; Maestri, T.; Martinazzo, M.; Masin, F.; Proietti Pelliccia, G.; Venafra, S.; Camy-Peyret, C. Demonstration of a physical inversion scheme for all-sky, day-night IASI observations and application to the analysis of the onset of the Antarctica ozone hole: Assessment of retrievals and consistency of forward modeling. *J. Quant. Spectrosc. Radiat. Transf.* **2024**, *329*, 109211. [CrossRef]
65. Carissimo, A.; De Feis, I.; Serio, C. The physical retrieval methodology for IASI: The δ-IASI code. *Environ. Model. Softw.* **2005**, *20*, 1111–1126. [CrossRef]
66. Liuzzi, G.; Masiello, G.; Serio, C.; Venafra, S.; Camy-Peyret, C. Physical inversion of the full IASI spectra: Assessment of atmospheric parameters retrievals, consistency of spectroscopy and forward modelling. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *182*, 128–157. [CrossRef]
67. Masiello, G.; Serio, C.; Maestri, T.; Martinazzo, M.; Masin, F.; Liuzzi, G.; Venafra, S. The new σ-IASI code for all sky radiative transfer calculations in the spectral range 10 to 2760 cm-1: σ-IASI/F2N. *J. Quant. Spectrosc. Radiat. Transf.* **2024**, *312*, 108814. [CrossRef]
68. Hilton, F.; Armante, R.; August, T.; Barnet, C.; Bouchard, A.; Camy-Peyret, C.; Capelle, V.; Clarisse, L.; Clerbaux, C.; Coheur, P.F.; et al. Hyperspectral Earth observation from IASI: Five years of accomplishments. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 347–370. [CrossRef]
69. Della Rocca, F.; De Feis, I.; Masiello, G.; Pasquariello, P.; Serio, C. Comparison of the IASI water deficit index and other vegetation indices: The case study of the intense 2022 drought over the Po Valley. In Proceedings of the Remote Sensing of Clouds and the Atmosphere XXVIII, Amsterdam, The Netherlands, 5–6 September 2023; Comerón, A., Kassianov, E.I., Schäfer, K., Picard, R.H., Weber, K., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2023; Volume 12730, p. 127300C. [CrossRef]
70. Ibrahim, I.; Abu Samah, A.; Fauzi, R.; Noor, N.M. The land surface temperature impact to land cover types. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B3*, 871–876. [CrossRef]
71. Ullah, S.; Qiao, X.; Abbas, M. Addressing the impact of land use land cover changes on land surface temperature using machine learning algorithms. *Sci. Rep.* **2024**, *14*, 18746. [CrossRef]
72. Hussain, S.; Raza, A.; Abdo, H.G.; Mubeen, M.; Tariq, A.; Nasim, W.; Majeed, M.; Almohamad, H.; Al Dughairi, A.A. Relation of land surface temperature with different vegetation indices using multi-temporal remote sensing data in Sahiwal region, Pakistan. *Geosci. Lett.* **2023**, *10*, 33. [CrossRef]

73. Dhole, A.; Kadaverugu, R.; Tomar, S.; Biniwale, R.; Sharma, A. Impact of land cover classes on surface temperature in the vicinity of urban lakes and vegetation patches: A non-parametric regression analysis over decadal data. *Earth Sci. Inform.* **2023**, *16*, 3947–3961. [CrossRef]
74. Kawashima, S.; Ishida, T.; Minomura, M.; Miwa, T. Relations between Surface Temperature and Air Temperature on a Local Scale during Winter Nights. *J. Appl. Meteorol.* **2000**, *39*, 1570–1579. [CrossRef]
75. Duveiller, G.; Hooker, J.; Cescatti, A. The mark of vegetation change on Earth's surface energy balance. *Nat. Commun.* **2018**, *9*, 679. [CrossRef] [PubMed]
76. Zhang, Y.; Dai, M. Analysis of the Cooling and Humidification Effect of Multi-Layered Vegetation Communities in Urban Parks and Its Impact. *Atmosphere* **2022**, *13*, 2045. [CrossRef]
77. Gao, Z.; Hou, Y.; Zaitchik, B.F.; Chen, Y.; Chen, W. A Two-Step Integrated MLP-GTWR Method to Estimate 1 km Land Surface Temperature with Complete Spatial Coverage in Humid, Cloudy Regions. *Remote Sens.* **2021**, *13*, 971. [CrossRef]
78. Wan, Z. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens. Environ.* **2014**, *140*, 36–45. [CrossRef]
79. Lu, L.; Zhang, T.; Wang, T.; Zhou, X. Evaluation of Collection-6 MODIS Land Surface Temperature Product Using Multi-Year Ground Measurements in an Arid Area of Northwest China. *Remote Sens.* **2018**, *10*, 1852. [CrossRef]
80. Yao, R.; Wang, S.; Wang, L.; Wei, J.; Junli, l.; Yu, D. A detailed comparison of MYD11 and MYD21 land surface temperature products in mainland China. *Int. J. Digit. Earth* **2020**, *13*, 1391–1407. [CrossRef]
81. Chance, K.; Kurucz, R. An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 1289–1295. [CrossRef]
82. Liou, K.N. *Radiation and Cloud Processes in the Atmosphere: Theory, Observation, and Modeling*; Oxford University Press: Oxford, UK, 1992. [CrossRef]
83. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice*; World Scientific: Singapore, 2000; Volume 2.
84. Baldridge, A.; Hook, S.; Grove, C.; Rivera, G. The ASTER spectral library version 2.0. *Remote Sens. Environ.* **2009**, *113*, 711–715. [CrossRef]
85. Masiello, G.; Serio, C.; Deleporte, T.; Herbin, H.; Behrendt, A.; Bosser, P.; Bock, O.; Wulfmeyer, V.; Pommier, M.; Flamant, C. Comparison of IASI water vapour products over complex terrain with COPS campaign data. *Meteorol. Z.* **2013**, *22*, 471–487. [CrossRef]
86. Serio, C.; Masiello, G.; Liuzzi, G. Demonstration of random projections applied to the retrieval problem of geophysical parameters from hyper-spectral infrared observations. *Appl. Opt.* **2016**, *55*, 6576–6587. [CrossRef] [PubMed]
87. Serio, C.; Masiello, G.; Camy-Peyret, C.; Liuzzi, G. CO2 spectroscopy and forward/inverse radiative transfer modelling in the thermal band using IASI spectra. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *222–223*, 65–83. [CrossRef]
88. De Feis, I.; Masiello, G.; Cersosimo, A. Optimal Interpolation for Infrared Products from Hyperspectral Satellite Imagers and Sounders. *Sensors* **2020**, *20*, 2352. [CrossRef]
89. Masiello, G.; Ripullone, F.; Feis, I.; Rita, A.; Saulino, L.; Pasquariello, P.; Cersosimo, A.; Venafra, S.; Serio, C. The IASI Water Deficit Index to Monitor Vegetation Stress and Early Drying in Summer Heatwaves: An Application to Southern Italy. *Land* **2022**, *11*, 1366. [CrossRef]
90. Serio, C.; Masiello, G.; Pasquariello, P.; Feis, I.D.; Mastro, P.; Falabella, F.; Cersosimo, A.; Venafra, S.; Pepe, A. Exploiting the IASI profiling capability for surface parameters, atmospheric temperature, and water vapour to design emissivity contrast and water deficit indexes to monitor forests' response to droughts and heatwaves. In Proceedings of the Remote Sensing of Clouds and the Atmosphere XXVII, Berlin, Germany, 5 September 2022; Comerón, A., Kassianov, E.I., Schäfer, K., Picard, R.H., Weber, K., Singh, U.N., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2022; Volume 12265, p. 1226502. [CrossRef]
91. Pasquariello, P.; Masiello, G.; Serio, C.; Mastro, P.; Liuzzi, G.; Della Rocca, F.; De Feis, I. Innovative remote-sensed thermodynamical indices to identify vegetation stress and surface dryness: Application to southern Italy over the last decade. In Proceedings of the Remote Sensing of Clouds and the Atmosphere XXVIII, Berlin, Germany, 5 September 2022; Comerón, A., Kassianov, E.I., Schäfer, K., Picard, R.H., Weber, K., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2023; Volume 12730, p. 127300D. [CrossRef]
92. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
93. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
94. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 1 February 2024).
95. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005. [CrossRef]
96. Breiman, L. Stacked Regressions. *Mach. Learn.* **1996**, *24*, 49–64. [CrossRef]
97. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York Inc.: New York, NY, USA, 2001.

98. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Dover Publications Inc.: New York, NY, USA, 1965.
99. Zhang, Y.; Li, J.; Liu, D. Spatial Downscaling of ERA5 Reanalysis Air Temperature Data Based on Stacking Ensemble Learning. *Sustainability* **2024**, *16*, 1934. [CrossRef]
100. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
101. van Prehn, J. Functions to Run XGBoost in MATLAB, 2019. Available online: https://it.mathworks.com/matlabcentral/fileexchange/75898-functions-to-run-xgboost-in-matlab (accessed on 15 November 2024).
102. Huang, G.B. Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Trans. Neural Netw.* **2003**, *14*, 274–281. [CrossRef]
103. Li, J.; Shen, Y.; Chen, L.; Ng, C.W.W. Rainfall Spatial Interpolation with Graph Neural Networks. In Proceedings of the International Conference on Database Systems for Advanced Applications, Tianjin, China, 17–20 April 2023; Wang, X., Sapino, M.L., Han, W.S., El Abbadi, A., Dobbie, G., Feng, Z., Shao, Y., Yin, H., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 175–191.
104. Montanari, A.; Nguyen, H.; Rubinetti, S.; Ceola, S.; Galelli, S.; Rubino, A.; Zanchettin, D. Why the 2022 Po River drought is the worst in the past two centuries. *Sci. Adv.* **2023**, *9*, eadg8304. [CrossRef]
105. Della Rocca, F.; De Feis, I.; Masiello, G.; Pasquariello, P.; Serio, C. Machine learning techniques for spatial interpolation of the IASI water deficit index. In Proceedings of the Remote Sensing of Clouds and the Atmosphere XXIX, Edinburgh, UK, 17–18 September 2024; Kassianov, E.I., Lolli, S., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2024; Volume 13193, p. 1319307. [CrossRef]