Validation of Regional-Scale Remote Sensing Products in China: From Site to Network

Shuguo Wang 1,2, Xin Li 2,3,4,*, Yong Ge 4,5, Rui Jin 2,3, Mingguo Ma 2,6, Qinhuo Liu 4,7, Jianguang Wen 7 and Shaomin Liu 8

1 School of Geography, Geomatics and Planning, Jiangsu Normal University, No. 101 Shanghai Road, Xuzhou 221116, China; swang@jsnu.edu.cn
2 Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, No. 320 Donggang West Road, Lanzhou 730000, China; jinrui@lzb.ac.cn
3 Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China
4 University of Chinese Academy of Sciences, Beijing 100049, China
5 State Key Laboratory of Resources & Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11A Datun Road, Beijing 100101, China; gey@lreis.ac.cn
6 School of Geographical Sciences, Southwest University, No. 2 Tiansheng Road, Chongqing 400715, China; mmg@swu.edu.cn
7 State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No. 20A Datun Road, Beijing 100101, China; liuqh@radi.ac.cn (Q.L.); wenjg@radi.ac.cn (J.W.)
8 State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, No. 19, Xinjiekouwai Street, Beijing 100875, China; smliu@bnu.edu.cn
* Correspondence: lixin@lzb.ac.cn; Tel.: +86-931-496-7249

Academic Editors: Richard Müller and Prasad S. Thenkabail
Received: 20 September 2016; Accepted: 21 November 2016; Published: 26 November 2016

Abstract: Validation is mandatory to quantify the reliability of remote sensing products (RSPs). However, this process is not straightforward and usually presents formidable challenges in terms of both theory and real-world operations. In this context, a dedicated validation initiative was launched in China, and we identified a validation strategy (VS). This overall VS focuses on validating regional-scale RSPs with a systematic site-to-network concept, consisting of four main components: (1) general guidelines and technical specifications to guide users in validating various land RSPs, particularly aiming to further develop in situ sampling schemes and scaling approaches to acquire ground truth at the pixel scale over heterogeneous surfaces; (2) sound site-based validation activities, conducted through multi-scale, multi-platform, and multi-source observations to experimentally examine and improve the first component; (3) a national validation network to allow for comprehensive assessment of RSPs from site or regional scales to the national scale across various zones; and (4) an operational RSP evaluation system to implement operational validation applications. Research progress on the development of these four components is described in this paper. Some representative research results, with respect to the development of sampling methods and site-based validation activities, are also highlighted. The development of this VS improves our understanding of validation issues, especially to facilitate validating RSPs over heterogeneous land surfaces both at the pixel scale level and the product level.

Keywords: validation; remote sensing product; technical specification; sampling; ground truth; pixel scale; heterogeneous surface
1. Introduction

In recent decades, unprecedented advances in characterising land surface parameters from satellite platforms have been achieved, boosting the development and a wide usage of numerous remote sensing products (RSPs) (e.g., [1–13]). Although these products provide reasonable spatial and temporal coverage, their accuracies and precisions should be independently assessed with reference data prior to utilisation. This process is well known as validation [14,15]. In practice, validation generally refers to assessing the uncertainties in remote-sensing-derived results (e.g., RSPs) via comparison with ground truth or inter-comparison with other alternative information that presumably represents the true state of a target.

Since validation is essential in various remote sensing applications, over the past two decades, huge efforts have been devoted internationally to address this issue in depth. These activities include the establishment of validation sites and networks as well as the contribution of large field campaigns and observation programmes to the topic of validation.

- Many core sites have been identified, such as the La Crau site for high-resolution SPOT data calibration/validation activities [16]; soil moisture (SM) observatories located in the Tibetan Plateau [17,18]; and numerous enhanced Earth Observing System (EOS) land validation core sites [19] recommended by the Committee on Earth Observation (CEOS) Land Product Validation (LPV) sub-group [20].
- Heavily instrumented observation networks have been established on scales ranging from regional to continental. These networks, such as the FLUXNET [21], the International Soil Moisture Network (ISMN) [22] and the Soil Climate Analysis Network (SCAN) [23], have paved the way towards meeting a broad range of national and international requirements for validation by continuously observing terrestrial variables.
- Intensive land surface/remote sensing experiments and field campaigns as well as research/observation plans/programmes can also greatly contribute to addressing validation issues, such as the Boreal Ecosystem–Atmosphere Study (BOREAS) [24], Cold Land Processes Field Experiment (CLPX) [25], the Complicate Observations and Multi-Parameter Land Information Constructions on Allied Telemetry Experiment (COMPLICATE) [26], Hydrological and Atmospheric Pilot Experiment in the Sahel (HAPEX) [27], First International Satellite Land Surface Climatology Project Field Experiment (FIFE) [28], Soil Moisture Experiment (SMEX) [29], Validation of the Land European Remote Sensing Instruments (VALERI) [30], Heihe Watershed Allied Telemetry Experimental Research (HiWATER) [31], the Global Climate Observing System (GCOS) [32], Critical Zone Observatory (CZO) [33], Terrestrial Environmental Observatories (TERENO) [34], Coordinated Enhanced Observing Period (CEOP) [35] and BigFoot [36].

It can be clearly seen that there has been a consensus to use diverse observatories converted from the site-to-network scale as a promising way to provide insight into the validation process. As a result, substantial outcomes, such as the establishment of a validation framework [19,37] and multi-scale sampling strategy in validation [38] as well as concrete validation activities concerning various land surface variables, were extensively reported (Table 1).

However, previous investigations have also demonstrated that at least two challenges remain in the validation process. First, in existing validation exercises, homogeneous land surfaces are a focus and are commonly selected as a validation field to deploy in situ measurements, while heterogeneous scenes are rarely discussed [39]. This type of validation scheme may be disputable since an intrinsic characteristic of land surface is heterogeneity, as opposed to homogeneity. Therefore, the validation approach and/or scheme fitted to homogeneous conditions may not be suitable to heterogeneous land surfaces. This could potentially lead to an unreliable assessment of the performance of an RSP. Second, although ground instrumentation is typically assumed to deliver reliable direct measurements on small scales, whether in situ measurements have reasonable spatial representativeness remains questionable. Biased estimation of ground truth at the pixel scale may be obtained due to inherent heterogeneity and the mismatch of scales between ground measurements and satellite footprint
This situation can notably worsen if point-based measurement is compared against RSPs with coarse resolution, particularly over extreme environments, e.g., cold and arid regions, where in situ measurements are sparsely distributed. Even when point-scale ground measurements are aggregated, the yielded pixel scale value may still not be reliable. This can be attributed to a lack of dense or better-distributed ground sampling, and, moreover, to deficiency in terms of scaling methods that are unable to account for the various characteristics of different RSPs and land surface heterogeneity [39,41]. As a result, this clearly limits the ability of obtaining trustworthy estimates of variables with strong spatial–temporal variability at the satellite pixel scale and often impedes the performance of direct validation between in situ measurements and RSPs.

The abovementioned research suggests that a framework to develop, design, and perform reasonable validation schemes and activities to acquire ground truth at the pixel scale over heterogeneous land surfaces is urgently needed. We must first ask, how do we integrate and get the utmost out of various ground observations collected at multiple scales to validate different RSPs, especially for those variables with strong heterogeneity? Second, can scale approaches be further developed to aggregate point-scale ground measurements to the pixel scale, and can such approaches improve spatial and temporal representativeness of heterogeneous land surfaces? These research questions are not yet fully understood. With the support of a Chinese national validation initiative, these questions, together with validation-related theories, methods, metrics, and scale effects, are intensively investigated.

The objective of this study is to present our research progress on the development of a validation strategy (VS) and some representative validation activities. The VS is composed of four main components, namely, general guidelines and technical specifications, site-based validation activities, a prototype of a Chinese national validation network, and an operational RSP validation system. Validation activities are conducted over reference sites dependent on multi-scale observations collected from multi-platform and multi-source experiments, such as HiWATER [31]. The remainder of the paper is organised as follows: Section 2 provides an introduction to the overall VS, which can be viewed as a general roadmap for the validation efforts we conceived and performed. In Section 3, the validation guidelines and technical specifications are elaborated in association with pertinent improvements in sampling design and scaling methods. Section 4 describes experimental validation exercises that have been undertaken over reference sites. Section 5 illustrates a networking validation attempt in China and an operational validation system before providing conclusions in the last section.

2. Validation Strategy

As mentioned earlier, validation is not a straightforward task. In contrast, in addition to inherent heterogeneity of land surfaces, validation is generally recognised as a challenging issue because RSPs may not be spatially and/or temporally consistent. In addition, as is well known, validation commonly comprises several key components, ranging from in situ measurements collection, modelling, and retrieval of land surface variables to scale-related analyses [15]. All these factors complicate the validation issue, calling for a better understanding of validation practice, including quantitative evaluation of uncertainties within the validation process.

Against this background, in 2011, the Ministry of Science and Technology of the People’s Republic of China (MOST) launched an initiative focused on the development and validation of regional-scale RSPs for terrestrial hydrological and ecological applications. The core mission of this programme is to derive long-term regional-scale land RSPs (listed in Table 1) and to investigate the validity of these developed products based on integrated airborne, satellite-based and ground-based measurements. Under this circumstance, validation-related theories, methods, evaluation scores, and scale effects can therefore be substantially studied.

Based on previous research achievements (e.g., [19,20,41–43]), an overall validation strategy (Figure 1) was identified. The VS aims at validating regional-scale RSPs with systematic site-to-network concept consisting of (1) general validation guidelines and technical specifications; (2) site-based field campaigns and intensive observations; (3) a national validation network; and (4) an operational validation system.
(1) As shown in Figure 1, validation guidelines and technical specifications constitute the first component of the VS, which can be regarded as a top-layer design to drive the whole VS. The validation guidelines provide universal principles and methodologies that can be applied to steer validation activities for diverse land RSPs, in which sampling design and scaling methods are stressed. In contrast, validation technical specifications are a set of dedicated implementation protocols towards 16 RSPs/variables of interest (as shown in Table 1). Each individual variable has a unique technical specification. Pertinent details about validation guidelines and technical specifications will be given in Section 3.2.

It is also worth mentioning that in this component, validation activity is essentially dependent on a priori knowledge of the reference site, such as its meteorology, topography, plant type, agricultural activity, and other site parameters. Then, we can better integrate and utilise diverse observations to further develop sampling design and scale methods. In this case, observations are mainly dependent on ground measurements acquired at conventional point scale or from a source of wireless sensor network (WSN) and footprint dimension as well as airborne remotely sensed imagery with high resolution (on the order of sub-meters and/or meters). Some research efforts related to this component will be highlighted in Section 3.3.

(2) The validity of proposed guidelines and specifications needs to be experimentally evaluated. Therefore, Component 2, i.e., field campaigns and intensive observations, is subsequently proposed in the VS. As depicted in Figure 1, activities in this component are to validate RSPs over well-established observation fields/sites, which links theories and methodologies to real-world validation exercises. On the one hand, in situ measurements (e.g., WSN) can be deployed based on a sampling scheme proposed in Component 1 to collect ground data. On the other hand, upscaling these measurements to the pixel scale value also relies on scaling approaches. As feedback, simultaneous satellite, airborne and ground-based remote sensing experiments (e.g., HiWATER) occurring at some key experimental areas (KEAs) can greatly contribute to inspect the validity and effectiveness of Component 1 by primarily validating relatively higher resolution (on the order of tens of meters) RSPs at site/local to watershed scale. Progress with respect to this aspect will be given in Section 4.

(3) However, readily available RSPs are usually presented at lower spatial resolution (of the order of kilometers to tens of kilometers) and with larger spatial coverage or even on a global scale basis. To validate this kind of products requires ground measurements to be collected across different climate zones, which depends on joint site-based validation efforts rather than merely over limited spatial extents/reference sites (e.g., within a river basin), as considered in Component 2. Such activities would make it possible to facilitate the establishment of a national validation network, as we accordingly proposed in Component 3.
In this context, at the prototype stage of this validation network, several anchor sites located at different climate zones and landscapes with sound observational facilities are initially selected. Synergetic multi-platform intensive observations or comprehensive remote sensing experiments are performed at these sites to discern and further rectify the weaknesses of the developed validation guidelines and specifications. Then, the amended Component 1 can be inspected by well-studied validation exercises again. Such a reiterative procedure is conducive to progressively improving and maturing the guidelines to make them more robust and standardised, with wider applicability and firm validity at the national scale. This provides possibilities to coordinate more validation test fields in the future, extending operational validation efforts to meet continental or global range application requirements. Details of this validation network can be found in Section 5.1.

(4) Technically, the above three components are integrated and realised by Component 4, i.e., an operational validation system. As a service portal, the functionalities and algorithms embedded in this system are designed to use multi-scale in situ and remote sensing data to validating quantitative RSPs at the regional or global scales. As shown in Figure 1, this system includes three parts related to data sets, accuracy evaluation and scale effect analysis. This setup facilitates the implementation of desired functions ranging from data host to product accuracy assessment. Being compliant with Component 1, scale effect can also be investigated in this process. The introduction of this system can be found in Section 5.2.

In conclusion, with commitment of continuous data collection over nationally representative sites, the conceived overall VS and involved work packages are expected to help us gain additional insight into validation issues and especially address the immediate scientific question of how to acquire representative ground truth in remote sensing validation (including modelling) over heterogeneous landscapes.

3. Validation Guidelines and Technical Specifications

This section describes the details of the first component of the VS, and presents the progress we have made pertinent to sampling design and scaling methods. However, before presenting these materials, the RSPs to be validated are identified.

3.1. Products of Interest

Currently, we focus on 16 land surface variables, as summarised in Table 1. For each variable, typical remote sensing observation methods and previous validation studies as well as our validation activities are presented. Based on discrepancies in spatial sampling and aggregation methods used in validation, the 16 variables can be grouped into the following two types (the first column in Table 1).

1. Continuous product. This type of variable features continuously varying values in time and space at each grid. Aggregating ground measurements into the satellite pixel scale can be performed using conventional spatial sampling methods (such as simple random sampling, systematic sampling, and stratified sampling). In addition, the aggregated value is a weighted sum of all samples, in which the sample weight is determined by the sampling method.

2. Categorical product. The value of each grid of this type of product is usually assigned as a group name instead of magnitudes and numbers in order or size. Pixel scale values can also be validated by ground-based spatial sampling, while the aggregation methods differ from those used for the continuous products.

Most of the variables in this table are usually observed on the ground at the point scale. However, different RSPs characterise varying spatial resolutions (e.g., soil moisture and LAI products), ground measurements are unable to cover all of these scales. Therefore, point-scale measurements (by fixed station or mobile device) must be rationally deployed with an appropriate sampling scheme. On this basis, the collected ground measurements can then be aggregated into the pixel scale to validate corresponding RSPs with the aid of upscaling approaches, particularly against coarse-scale products (e.g., [44]).
Table 1. Land surface variables to be validated. The RS observation method refers to the commonly used satellite-based technique.

<table>
<thead>
<tr>
<th>Types of RSPs</th>
<th>Variables</th>
<th>RS Observation Method</th>
<th>Spatial Resolution</th>
<th>Previous Studies</th>
<th>Our Validation Efforts (Basic Information and References)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous product</td>
<td>Albedo</td>
<td>VNIR *</td>
<td>30 m–1 km</td>
<td>[39,45–52]</td>
<td>Multi-scale validation strategy for satellite albedo products [53]; Optimal nodes selectiveness from WSN to fit field scale albedo observation and validation [54]</td>
</tr>
<tr>
<td></td>
<td>Emissivity</td>
<td>TIR</td>
<td>90 m–1 km</td>
<td>[11,55–60]</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>ET</td>
<td>VNIR, TIR</td>
<td>30 m–1 km</td>
<td>[61–65]</td>
<td>Measurements of ET from EC and LAS systems [66] and validation of remotely sensed ET [67] in the Hai River Basin; A flux observation matrix was established over the HRB [68]; Estimating and validating soil evaporation and crop transpiration [69,70]; Upscaling ET measurements from the field to larger scales [71,72]; A framework for validating remotely sensed ET [73]; Characterising the footprint of EC and LAS measurements to validate satellite-based surface fluxes [74]</td>
</tr>
<tr>
<td></td>
<td>FAPAR</td>
<td>VNIR</td>
<td>30 m–1 km</td>
<td>[75–80]</td>
<td>Scaling of remotely sensed and measured FAPAR [81,82]</td>
</tr>
<tr>
<td></td>
<td>FCover</td>
<td>VNIR</td>
<td>30 m–1 km</td>
<td>[75–78]</td>
<td>Evaluation of sampling methods for validation of remotely sensed FCover [83]</td>
</tr>
<tr>
<td></td>
<td>LAI</td>
<td>VNIR, LiDAR</td>
<td>30 m–1 km</td>
<td>[43,75–78,84,85]</td>
<td>Assessment of the MODIS LAI product in the meadow steppe [86]; Sampling strategy for observing and validating remotely sensed LAI product over heterogeneous land surface [87,88]; Development of a WSN used for validating LAI products [89]; Upscaling algorithm to obtain ground truth of LAI time series over heterogeneous land surface [90]</td>
</tr>
<tr>
<td></td>
<td>LST</td>
<td>TIR</td>
<td>90 m–1 km</td>
<td>[11,55,58,60,91–93]</td>
<td>Evaluation of LST retrieval from FY-3B/VIRR data [94]; Analysis of scale mismatch between in situ and remote sensing LST product [95]; Evaluation and validation of the MODIS LST in arid regions of China [96,97]; Validation of LST estimated from ASTER data in the HRB [98]</td>
</tr>
<tr>
<td></td>
<td>Net radiation</td>
<td>VNIR, TIR</td>
<td>30 m–1 km</td>
<td>[99–103]</td>
<td>Analysis of representativeness errors of point-scale ground-based solar radiation measurements in the validation of RSPs [104]; A framework to obtain high resolution surface radiation products [105]</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Types of RSPs</th>
<th>Variables</th>
<th>RS Observation Method</th>
<th>Spatial Resolution</th>
<th>Previous Studies</th>
<th>Our Validation Efforts (Basic Information and References)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous product</td>
<td>NPP</td>
<td>VNIR, LiDAR</td>
<td>30 m–1 km</td>
<td>Validation of MODIS GPP product at ten flux sites in northern China [111]; Spatial sampling design for estimating regional GPP [112]; Analysis of spatial representativeness and uncertainty of flux measurement for upscaling ecosystem productivity to field scale [113]</td>
<td>Validation of MODIS GPP product at ten flux sites in northern China [111]; Spatial sampling design for estimating regional GPP [112]; Analysis of spatial representativeness and uncertainty of flux measurement for upscaling ecosystem productivity to field scale [113]</td>
</tr>
<tr>
<td>Snow cover area</td>
<td></td>
<td>VNIR</td>
<td>30 m–1 km</td>
<td>Analysis of snow cover variations and controlling factors at upper HRB [124]; Improving mountainous snow cover fraction mapping [125]; Downscaling snow cover fraction data in mountainous regions [126]</td>
<td>Analysis of snow cover variations and controlling factors at upper HRB [124]; Improving mountainous snow cover fraction mapping [125]; Downscaling snow cover fraction data in mountainous regions [126]</td>
</tr>
<tr>
<td>Soil moisture</td>
<td></td>
<td>Active/passive microwave</td>
<td>1–25 km</td>
<td>Analysis of temporal variability of uncertainty in pixel-wise soil moisture [137]; Design and establishment of WSN for capturing soil moisture over heterogeneous land surface [138,139]; Upscaling in situ soil moisture observations to pixel averages [140–142]</td>
<td>Analysis of temporal variability of uncertainty in pixel-wise soil moisture [137]; Design and establishment of WSN for capturing soil moisture over heterogeneous land surface [138,139]; Upscaling in situ soil moisture observations to pixel averages [140–142]</td>
</tr>
<tr>
<td>Vegetation index</td>
<td></td>
<td>VNIR</td>
<td>30 m–1 km</td>
<td>Validation of the MODIS NDVI products in the HRB [148]; Analysis of temporal dynamics of spatial heterogeneity of NDVI [149]</td>
<td>Validation of the MODIS NDVI products in the HRB [148]; Analysis of temporal dynamics of spatial heterogeneity of NDVI [149]</td>
</tr>
<tr>
<td>Categorical product</td>
<td>LUCC</td>
<td>VNIR</td>
<td>30 m–1 km</td>
<td>LUCC in the middle reaches of the HRB over 2000-2011 [155]; Finer resolution land-cover mapping in the HRB [156]</td>
<td>LUCC in the middle reaches of the HRB over 2000-2011 [155]; Finer resolution land-cover mapping in the HRB [156]</td>
</tr>
</tbody>
</table>

* VNIR: visible and near infrared; TIR: thermal infrared; ET: evapotranspiration; EC: eddy covariance; LAS: large aperture scintillometer; HRB: Heihe River Basin; FAPAR: fraction of photosynthetically active radiation; FY: Fengyun meteorological satellite; LAI: leaf area index; LiDAR: light detection and ranging; LST: land surface temperature; NPP: net primary product; GPP: gross primary product; NDVI: normalised difference vegetation index; and LUCC: land use and land cover change.
3.2. Validation Guidelines and Technical Specifications

3.2.1. Validation Guidelines

Validation guidelines and technical specifications are the cornerstones of the validation framework [161]. The latter focuses on detailed validation protocols towards each specific land surface RSPs of interest, as described in Section 3.2.2. In contrast, the former is a baseline document that provides general considerations and basic procedures/approaches used in validating various land surface RSPs, and the content is roughly shown as follows:

- Validation criteria include accuracy, precision, completeness, temporal variation, spatial pattern, spatial and temporal consistency.
- Validation metrics, such as bias, mean absolute error (MAE), root mean square error (RMSE), mean absolute percent error (MAPE) and correlation coefficient R.
- Basic procedures and approaches used to validate either continuous or categorical RSPs are provided at the pixel scale level and at the product level, as illustrated in Figure 2.

- Basic procedures and approaches used to validate continuous RSPs at the pixel scale level. In this aspect, direct validation can capitalise on in situ measurements collected at the point scale and/or footprint scale (e.g., EC system). Because both the location and number of sampling points can directly affect the reliability of pixel scale representative estimates, ground sampling design is conducted based on variables characteristics with the aid of geostatistical methods, such as the mean of the surface with non-homogeneity (MSN) [162,163]. In the upscaling process, ground truth at the pixel scale is derived through integrating in situ measurements, such as various kriging models. Then, the upscaled ground measurements can be used to compare against RSP.

- Basic procedures and approaches used to validate categorical land RSPs at the pixel scale level are similar to the ones validating continuous RSPs. Ground sampling design also plays an important role in this process. For instance, in direct validation, ground measurements acquired at sampling points can be used to compare against remote sensing estimates.

- Basic procedures and approaches used to validate land RSPs at the product level. The first is to select typical sites, pixels and time periods across different biomes and landscapes. Next, in temporal, product accuracy should be assessed via comparison against time series reference data at the pixel scale over each specific site. On this basis, in spatial, product accuracy should be further assessed over a widely distributed set of locations. Finally, overall evaluation of an RSP can be achieved by combining the assessed results acquired at both time and space. The spatial and temporal consistency and uncertainty of a product can then be well evaluated and quantified via independent measurements over wildly representative locations and time periods.

- General guidance on field site selection and corresponding ground instrumentation. This aspect suggests detailed criteria on how to choose or build a proper validation site and how to deploy various instruments to acquire high-quality ground measurements. Primarily, a site should be spatially representative of a given biome/ecosystem, considering national physiographic conditions. Basic observation facilities are also needed, including meteorology (e.g., precipitation and air temperature), eco-hydrological elements (e.g., soil temperature and moisture) and flux measurements. Other selection principles should be considered, e.g., accessibility, site extent, ambient environment, and commitment to long-term scientific study.
Validation Technical Specifications

As shown in Table 1, it is well known that different RSPs can be preferably derived using different remote sensing methods, such as LAI or soil moisture, which are observed via optical or microwave remote sensing. In this case, related validation efforts such as ground sampling design, site selection, site extent, validation method (direct/indirect/inter-comparison) and specific evaluation metrics to be used are varied. Compared to the validation guidelines providing general conditions and universal methodologies, technical specifications of validation are more specific and technical. These specifications refer to a set of procedures/protocols that benchmark for implementing validation exercises towards 16 RSPs/variables of interest (as shown in Table 1). Each individual RSP/variable has one dedicated technical specification. Here, we give an example of validation technical specification of ET. The content is as follows:

- Validation criteria are data completeness, product accuracy, spatial and temporal variation, spatial and temporal consistency.
- Validation metrics are bias, RMSE, MAPE, mean relative error (MRE) and R.
- Procedures and approaches used in direct validation include the following: The first step is the acquisition of ground truth. At the watershed/regional scale, based on the measurements of water budget components (e.g., precipitation, runoff), water balance equation can be used to calculate yearly averaged ET result as ground truth. At small (point-, footprint- and pixel-) scales, direct flux measurements from Lysimeter, EC system and LAS, etc. are recommended. In the second step, at the river basin/watershed scale, spatial and temporal characteristics of long-term ET estimates are evaluated. At small scales, the accuracy of remotely sensed ET with instantaneous and daily/monthly averaged values as well as temporal variability are evaluated. Error sources and uncertainties are quantitatively assessed.
- Procedures and approaches used in indirect validation are the following: The goal of indirect validation is to validate ET estimates derived from low resolution (e.g., MODIS, NOAA/AVHRR pixel level) satellite observations. This involves using ground-based measurements and multi-source remote sensing information with different levels of spatial resolution, which is capitalised on a step-by-step concept, ranging from high resolution (e.g., airborne mission), medium resolution (e.g., Landsat, ASTER pixel level) to low spatial resolution, considering scaling effect correction in the upscaling process.
- Procedures and approaches used in inter-comparison are the following: Given multiple ET products, well-studied products can be used to evaluate others in need of validation. Evaluation
process biases are observed when assessing product accuracy and consistency, spatial and temporal characteristics and regional applicability. In terms of multiple ET estimation models, the same input data (e.g., RS observations and forcing data) are used to drive different ET estimation models. Apart from assessing the accuracy and spatial-temporal characteristic of derived results, the validation process also compares the impact caused by diverse model structures, mechanisms and parameterization schemes, etc., to acquire knowledge regarding the applicability, limitation, and uncertainty of different models.

- Ground instruments are recommended, such as Lysimeter, EC system, and LAS as well as related measuring protocols and data processing instructions.
- Finally, an evaluation report should be generated with respect to the above validation efforts and product performance.

3.3. Progress on the Development of Sampling Design and Scale Approach

Obtaining ground truth at the pixel scale requires a well-designed sampling scheme, high quality data sets and rational statistical inference of measured samples. For homogeneous pixels, a few samples are adequate for capturing the dominant information for a pixel grid. If a heterogeneous pixel is presented, additional samples are usually desired to reach an unbiased estimation with a satisfactory accuracy [161]. Hence, how to use a limited number of sampling points to optimise their spatial distribution and therefore acquire reliable ground truth at the pixel scale is a question to be answered. This topic is investigated based on the characteristics of land surface, e.g., spatial autocorrelation and homogeneity. In general, if there is a lack of spatial autocorrelation of ground objects, classical sampling methods (e.g., simple random sampling, systematic sampling, and stratified sampling) can be used. Otherwise, more complicated spatial sampling methods, such as the MSN spatial sampling optimisation scheme [162,163], are recommended. The following section focuses on the progress of the development of the sampling design and upscaling approach over heterogeneous land surfaces, as this issue is considered in Component 1 in Figure 1.

3.3.1. Sampling Design

Wang et al. [112] presented a geostatistical-model-based method to optimise the spatial sampling design for estimating GPP. This investigation occurred in the Babao River Basin, located upstream of the Heihe River Basin (HRB) over grassland. Sampling locations were optimised based on a stratified block kriging (StrBK) approach, taking the heterogeneity (represented by stratification and anisotropy) into account. The error variance of regional GPP estimation decreased by 10.1% compared with a sampling scheme without considering land surface heterogeneity. Over the same watershed, Ge et al. [164] reported a universal co-kriging (UCK) model-based sampling design optimisation. This scheme was embodied in the design of a WSN that monitored three eco-hydrological variables (LST, precipitation, and soil moisture). The results demonstrated that the proposed sampling method could consider the relationship of target variables and environmental covariates as well as other spatial statistics. Compared with a sampling scheme without consideration of the multivariate correlation, the proposed design performed better by reducing prediction error variance. Another study occurred in the midstream of the HRB, as reported in [88], an efficient and robust LAI sampling strategy was developed based on multi-temporal prior knowledge (SMP) for long-term, fixed-position LAI observations. The proposed SMP-based sampling scheme was compared with another four methods, including random sampling, systematic sampling, sampling based on the land-cover map, and sampling based on vegetation index prior knowledge, using the PROSAIL model-based simulation analysis. The results indicated that the average RMSE of the LAI reference maps decreased from 0.12 to 0.05, and the relative error can be reduced from 6.1% to 2.2%.

3.3.2. Scaling Methods

From the perspective of the upscaling issue, variables with strong heterogeneity (e.g., ET and soil moisture) were intensively investigated. Ge et al. [71] proposed an area-to-area regression
The kriging (ATARK) method to upscale sensible heat flux observations from EC scale to LAS footprints. Observations sourced from the footprints of 17 EC and four LAS systems were analysed. The results revealed the proposed ATARK was capable of delivering accurate predictions, with a high $R^2$ value ($\geq 0.9$) acquired for the comparison between upscaled EC and LAS measured flux. Furthermore, Liu et al. [72] compared five upscaling methods and developed a combined method to acquire ET ground truth at the satellite pixel scale. Based on multi-site measurements from 11 EC systems and four groups of LASs in the HRB, the authors indicated that three simple upscaling methods—namely, the arithmetic average method, area-weighted method, and footprint-weighted method—showed promise when used over homogenous surfaces, with an average error of approximately 6%. However, for heterogeneous landscapes, auxiliary variables should be introduced to characterise heterogeneity, such as the integrated Priestley-Taylor equation method or the ATARK method, and can be used to improve the upscaled ET results. On this basis, a combined method was proposed to acquire both instantaneous and daily ET ground truth at the satellite pixel scale with the moment of MODIS overpass.

Kang et al. [140] analysed regression kriging (RK)-based upscaling scheme for soil moisture observed via WSN with three types of remote sensing information. Compared with the ordinary kriging method, the derived spatial inferences showed that the RK method was able to decrease the RMSE by approximately 5%. Wang et al. [141] proposed a geostatistical approach to upscale soil moisture observations with unequal precisions. This approach considered random measurement errors and a Monte Carlo simulation in association with a block kriging (BK) upscaling strategy. When comparing the three upscaling approaches (simple average, error-free BK, and error-perturbed BK), the results showed that the aggregated soil moisture estimates were comparable, while the error-perturbed BK approach outperformed, with a minimum error standard deviation ($<0.01 \text{ m}^3 \text{ m}^{-3}$).

In addition, as shown in Component 1 in Figure 1, airborne mission is included. Using the observations acquired from a compact airborne spectrographic imager (CASI) aboard aircraft, a land-cover-based Linear Bi-directional Reflectance Distribution Function (BRDF) unmixing (LLBU) algorithm was proposed to estimate albedo over farmland in the HRB. The retrieved CASI albedo with 5 m resolution showed promising high accuracy, with an RMSE of 0.013 compared against in situ measurements. In addition, good performance was observed when the CASI albedo was upscaled to a 500 m pixel scale, with an RMSE of 0.019 compared with MODIS products [165].

The abovementioned research progress lays the foundation for answering several scientific questions; i.e., how could scale approaches be further developed to aggregate point-scale ground measurements to the satellite pixel scale, and how could the generated values improve spatial and temporal representativeness of heterogeneous land surfaces? In this case, there is an article [166] summarising our various upscaling research.

4. Validation Activities at Reference Sites

As shown in previously reviewed studies, a key component of a validation framework has been substantial short-term and large-scale intensive field campaigns. In addition, in our VS, to experimentally evaluate the validity of Component 1 in Figure 1 requires considerable validation efforts, with various measurements jointly collected through multi-source and multi-platform venues. In this section, we focus on site-based validation activities (presented as Component 2 in Figure 1) in a twofold manner: validation efforts in HiWATER and validation experiments over other places.

4.1. Validation Efforts in HiWATER

HiWATER was a watershed-scale remote sensing experiment conducted in the HRB in Northwest China [31]. It greatly supported validation studies. This is because first, the HRB has a unique natural environment, which spans the mountain cryosphere (upstream), coexisting forest, irrigated oasis and farmland (midstream), and natural oasis, semi-desert, and desert areas (downstream). Such diverse landscapes allow us to validate different ecological and hydrological RSPs. Second, numerous experimental activities that concern remote sensing and land surface processes have been
implemented in this river basin, resulting in robust research achievements and datasets as well as sound observation facilities [167]. Finally, remote sensing validation and its related issues were highly emphasised in HiWATER, reflected by the implementation of specific experimental configuration and observatory deployment that can contribute to validation activities.

In HiWATER, ground truth is not limited to being collected at the point scale; there is also the pixel scale based on WSN and the footprint scale based on EC system and LAS measurements.

Two ground-based WSNs were established over heterogeneous farmland in the middle stream of the HRB in 2012. One WSN was an eco-hydrological WSN (EHWSN) designed to capture spatial-temporal variations of soil moisture, soil temperature, and LST (Figure 3) [138,139]. It is worth mentioning in Figure 3 that a flux observation matrix (the oblique red domain) was identified in HiWATER. Because this area is heterogeneous, a 4 km × 4 km site was further identified (blue box) that can then be used to validate various MODIS products to minimise the PSF effect and geo-location uncertainties of sensors. The other WSN was LAINet, with each node equipped with light sensors to provide continuous measurements on corn LAI in growing season [89]. Such ground-based facilities can provide a priori spatial distribution information on eco-hydrological variables and can be used to validate corresponding RSPs on the order of a kilometer, i.e., approximately the MODIS pixel size. The utilisation of these WSNs in validation is highlighted by several case studies, as described in more detail below.

Figure 3. The setup of the EHWSN deployed in the Yingke and Daman irrigation districts in the middle stream of the HRB. The left panel shows the location, KEAs, and validation sites identified of the HRB. The right panel shows the EHWSN.

(1) Validating LST and soil moisture using the EHWSN. Yu and Ma [95], investigated the scale mismatch between in situ and remotely sensed LST. Two types of ground-based observations were obtained from a CNR4 net radiometer and an SI-111 infrared radiometer. Then, these two ground-based measurements were compared with MODIS LST products (MOD11A1) with 1 km resolution, while an LST retrieval of 3 m resolution derived from airborne thermal-infrared observations was used to assess land heterogeneity. Based on a semi-variance analysis, the results suggested that it was not reliable to directly use ground-based LSTs to represent their located MODIS pixel value. Surface heterogeneity triggered strong spatial variations of LST and pronounced errors in the validation process. Han et al. [168] estimated soil moisture at the footprint scale using a novel Cosmic-ray Soil
Moisture Observing System (COSMOS) over heterogeneous farmland, and in situ measurements acquired by the EHWSN were used to calibrate and validate soil moisture estimates.

(2) Validating vegetation index products using LAINet. The primary objective of LAINet was to collect continuous corn LAI measurements using 42 in situ WSN nodes. Thus, a comparative analysis was conducted among different ground measurements and MODIS observations [89]. This study included a direct comparison between LAINet measurements and data collected by LAI-2000. Daily and 5-day integrated times series analysis and aggregated ground LAINet LAI values were compared with satellite remotely sensed data. In [148], an evaluation of the MODIS NDVI products (MOD09GQ and MYD09GQ) was performed across six land use types over nearly one complete growing season. The spatial heterogeneity and scale effects of the NDVI were investigated. The study reported that spatial heterogeneity was commonly found in MODIS data at a resolution of 250 m for different land types. Upscaling of in situ NDVI data based on high spatial resolution satellite imagery can noticeably enhance the validation accuracy of NDVI estimates at the pixel level.

In addition to the WSNs, a thematic experiment within HiWATER was the multi-scale observation experiment on evapotranspiration (MUSOEXE), which established a flux observation matrix in the middle reaches of the HRB from May to September 2012. This observation matrix was composed of 22 EC systems, eight LAS systems and 21 automatic meteorological stations (AMS) that were used to deliver footprint-scale measurements (Figure 3) [68,72]. Multi-scale (e.g., on the order of 100 m to 1–2 km) observations of meteorological elements and land surface parameters were collected for validating regional ET over heterogeneous surfaces. In addition, these comprehensive measurements were beneficial for identifying scale effects and providing ground truth that coincides with the development of the remote sensing models and scaling approaches for validation purposes. As demonstrated in [69,70], Song et al. validated corn transpiration and soil evaporation estimated by the two-source energy balance model (TSEB) at site and field scales based on the EC and stable oxygen and hydrogen isotope measurements. In [73], an innovative validation framework including quantification of the spatial heterogeneity, optimisation of the ground sampling strategy, multi-scale measurements, upscaling theory, uncertainty analyses, and validation method (direct validation, indirect validation and cross validation) was proposed to validate ET RSPs at different scales. In conclusion, as a comprehensive remote sensing experiment, HiWATER felicitously connects validation theories with detailed validation exercises.

4.2. Other Validation Experiments

Validation activities were also performed in study sites other than HiWATER. A typical study included observing fluxes using EC and LAS systems to validate the ET product in the Hai River Basin, across typical underlying surfaces of the northern mountains (Miyun), croplands in the central suburbs (Daxing) and croplands in the southern plains (Guantao) [66,67]. Jia et al. [67] proposed an innovative validation method based on multi-source ET measurements. The validation results included accuracy assessment, analysis of error source and uncertainty within the validation process at the basin and local scales. Both scales showed a good agreement between the estimated and observed ET. Bai et al. [74] investigated the effect of footprint characteristics for validating satellite-based surface fluxes in two river basins. The results showed that the footprint is crucial in defining a consistent spatial scale between ground measurements and satellite-based surface flux estimates in validation, particularly for heterogeneous surfaces and high-resolution remote sensing data.

A study of accuracy assessment of multi-scale LAI observations occurred in a meadow steppe of Hulunber, Inner Mongolia, China [86]. Datasets involved were ground-based measurements, MODIS C5 LAI and land cover type products with 1 km and 500 m resolution, respectively, as well as LAI maps derived from Environment and Disaster Monitoring Small Satellite Constellation of China (HJ)-1A/1B CCD images with 30 m resolution. A slight overestimation in the averaged magnitude of 0.4 was found for the MODIS LAI product compared with the HJ-retrieved LAI maps, and the relative absolute errors of the product ranged from 10% to 50%. Ding et al. [149] introduced a concept
of mean length variability to be used as an index to quantify spatial heterogeneity over croplands for multi-temporal NDVI, near-infrared, and red reflectance. This investigation suggested that the spatial heterogeneity varied with the changes in the fraction of vegetation cover, and a spatial resolution larger than 120 m could effectively limit the difference of spatial heterogeneity among different remote sensing observation methods.

In addition to these individual activities, other field campaigns and intensive observations have been performed, particularly at anchor sites within a national validation network being coordinated (see Section 5). As a result, we are highly confident that the validation experience learned from local-scale validation is rewarding to leverage validation exercises to be practiced at larger scales.

5. A National Validation Network and an Operational Validation System

This Section describes the last two components proposed in the overall VS. To some extent, the role of these two parts can be regarded as a further operational usage stage beyond the first two components in Figure 1.

5.1. National Network for Validating RSP

As stated earlier, multi-platform and multi-scale validation activities allow for systematically verifying and improving the proposed validation guidelines and technical specifications. In the meantime, uncertainties in various RSPs can be quantitatively evaluated using diverse validation metrics. Experience obtained from such validation exercises is helpful for establishing a Chinese national validation network. Such a network is being coordinated as a prototype by initially selecting typical stations/sites with sound observational foundations. Currently, 12 anchor stations are chosen over seven representative regions in China [169]. The distribution of these sites as well as each dominant landscape is illustrated in Figure 4. In addition, four of these stations—namely, the Huailai Station, the Hulunber Station, the Jingyuetan Station, and the Heihe Station—have been selected as core observation sites. Basic information about the former three stations is described in Table 2. For the most heavily instrumented Heihe Station, a very detailed description can be found in [169].

Figure 4. Sketch map of anchor sites in the validation network [169].
Over the past few years, various validation experiments, including dedicated ground-based remote sensing experiments and intensive field campaigns, were individually and/or jointly conducted. Datasets are collected at these four sites for investigating remote sensing and validation mechanisms and theories. Because each station features distinct local landscapes, specific validation target variables are therefore identified [169]. Validation activities over such diverse environments can still help further improve the validation guidelines and specifications developed in Component 1. This procedure can be progressively performed until it fulfils the validation requirement over various surface conditions and multiple scales. Meanwhile, principles on the establishment of observation fields, deployment of instrumentation, data processing procedures, data quality control, data sharing and product evaluation metrics, etc. can be standardised. Accordingly, more sites and stations can be united to form a recognisable nationwide remote sensing validation network.

Table 2. Summary information of core observation stations in the validation network [169].

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Core Site and Size</th>
<th>Location</th>
<th>Meteorology</th>
<th>Land Cover</th>
<th>Scientific Aims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huailai Station</td>
<td>East garden controlled</td>
<td>Prov.: Hebei;</td>
<td>Precip.: 650; Temp.:</td>
<td>Type: Cropland;</td>
<td>Remote sensing mechanism; scale approach</td>
</tr>
<tr>
<td></td>
<td>experimental site 100 m</td>
<td>Long.: 115.783;</td>
<td>10</td>
<td>LAI: 6.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>× 100 m</td>
<td>Lat.: 40.349; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>East garden 1.5 km × 2</td>
<td>Prov.: Hebei;</td>
<td>Precip.: 650; Temp.:</td>
<td>Type: Cropland;</td>
<td>Remote sensing mechanism; ecological protection; crop growth simulation model</td>
</tr>
<tr>
<td></td>
<td>km</td>
<td>Long.: 115.783;</td>
<td>10</td>
<td>LAI: 6.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lat.: 40.349; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Guanting reservoir 2 km</td>
<td>Prov.: Hebei;</td>
<td>Precip.: 650; Temp.:</td>
<td>Type: Water;</td>
<td>Ecological protection; water resource utilization</td>
</tr>
<tr>
<td></td>
<td>× 2 km</td>
<td>Long.: 115.740;</td>
<td>9</td>
<td>LAI: 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lat.: 40.352; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>475</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jingyuetan Station</td>
<td>Taipingpu controlled</td>
<td>Prov.: Jilin;</td>
<td>Precip.: 520; Temp.:</td>
<td>Type: Cropland;</td>
<td>Radiative transfer theory; soil freeze/thaw process; scale approach</td>
</tr>
<tr>
<td></td>
<td>experimental site 200 m</td>
<td>Long.: 125.401;</td>
<td>4.4</td>
<td>LAI: 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>× 200 m</td>
<td>Lat.: 43.998; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>211</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dehui corn 2 km × 2 km</td>
<td>Prov.: Jilin;</td>
<td>Precip.: 604; Temp.:</td>
<td>Type: Cropland;</td>
<td>Validation; water balance; crop simulation growth model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long.: 125.360;</td>
<td>4.8</td>
<td>LAI: 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lat.: 44.119; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>207</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hulunber Station</td>
<td>Stipa baicalensis site</td>
<td>Prov.: Inner Mongolia;</td>
<td>Precip.: 380; Temp.:</td>
<td>Type: Temperate</td>
<td>Validation; grassland ecosystem process model</td>
</tr>
<tr>
<td></td>
<td>3 km × 3 km</td>
<td>Long.: 120.087;</td>
<td>−1.0</td>
<td>meadow steppe;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lat.: 49.342; Elevation:</td>
<td></td>
<td>LAI: 2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>642</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cropland site 3 km × 3</td>
<td>Prov.: Inner Mongolia;</td>
<td>Precip.: 380; Temp.:</td>
<td>Type: Cropland;</td>
<td>Validation; crop simulation growth model</td>
</tr>
<tr>
<td></td>
<td>km</td>
<td>Long.: 120.034;</td>
<td>−1.0</td>
<td>LAI: 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lat.: 49.308; Elevation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>638</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Precip.: Annual averaged precipitation (mm); Temp.: Annual averaged temperature (°C); Prov.: Province; Long.: Longitude (°E); Lat.: Latitude (°N); Ele.: Elevation (m).

5.2. An Operational RSP Validation System

As the last Component in Figure 1, a web-based land surface remote sensing product validation system (LAPVAS) is designed to use multi-scale in situ and remote sensing data for validating RSPs at regional and/or global scales. This system facilitates functionalities in data host and validation execution. The latter is performed based on single-point and multi-location comparisons in the case of homogenous and heterogeneous land surfaces, respectively. As shown in Figure 5, this system includes three key elements: (1) A validation database subsystem contains an original database that hosts ground measurements and, a validation data repository that stores the pixel scale reference data. Various analysis tools used to generate spatial-temporal scale matched validation data from
the original database are also included in this subsystem; (2) an RSP accuracy evaluation subsystem to quantify the uncertainties and consistency of a product, with accuracy, precision, completeness, spatial and temporal consistency, etc. are used as validation criteria; (3) an inter-external interface exchanges information among the validation database subsystem, the evaluation subsystem, and the RSP data flow.

![Figure 5. The structure and workflow of the validation system of the LAPVAS.](image)

The original data in the validation database subsystem are collected from in situ measurements and/or from well-qualified high-resolution products for indirect validation purposes. The preprocessing of spatial scale transformation and temporal-scale matching are implemented to generate spatial-temporal matches among validation datasets and RSPs to be validated. The candidate RSP is preferred to be spatially and temporally consistent with reference data, which is processed by the inter-external interface to generate comparable data in the matched table (i.e., the data from candidate RSP and the counterparts from the validation data). Subsequently, the evaluation subsystem provides three methods in validating: direct validation, indirect validation and inter-comparison, according to the land cover heterogeneity. Finally, an evaluation report can provide validation results, particularly the data information, validation method, and accuracy of the RSP being validated.

The above three parts of LAPVAS are built using various essential service modules, such as data-read service modules, data-insert service modules, data-associated service modules and
evaluation-analysis service modules, packaged by a standard web-service description language. Users can flexibly choose these service modules and combine them through the user interface to form a user-defined programme according to validation requirements. Technically, a framework called a service-oriented architecture (SOA) is adopted in LAPVAS. One promising feature of this SOA is the ability to modularise LAPVAS using low cohesion and loose coupling. In this sense, the modularised system can be managed in an operational way, which places less importance on the sequence of service modules in LAPVAS.

6. Conclusions

In this study, we introduce an overall validation strategy with a site-to-network concept. The VS comprises four main components: (1) general guidelines and technical specifications that can guide users to validate various land RSPs; (2) substantial validation activities to experimentally examine and improve the first Component; (3) a national validation network to allow comprehensive assessment of RSPs from site or regional scales to the national scale across various zones; and (4) an RSP evaluation system to implement operational validation applications. The proposed VS systematically focuses on various aspects of validation issues, including its measures, approaches, and real-world application.

Validation guidelines and a series of technical specifications are proposed. The former provides general conditions and universal methodologies to validate either continuous or categorical RSPs both at the pixel scale level and at the product level. Validation technical specifications focus on validation protocols towards 16 land surface RSPs of interest. Research outcomes related to ground sampling design and scaling method demonstrated that these investigations have been beneficial in quantitatively calculating various statistical indicators for evaluating RSPs, optimising the spatial layout of sampling points, and obtaining unbiased and representative aggregated ground truth data at the pixel scale.

Substantial intensive observations and field campaigns performed from site to local to regional scales greatly supported validation activities. The HiWATER experiment occurred in the HRB, as well as other experiments conducted in China, these studies successfully collected high-quality datasets at multiple scales using WSNs, field measurements with footprint dimensions, airborne remote sensing, and other intensive in situ measurements. The utilised sampling schemes and deployment of ground instrumentation, etc., are compliant with the first component of the VS. In this context, considerable research results have already been yielded, including validation efforts addressed soil moisture, LST, ET and vegetation index, etc.

In addition, within this strategy, we are coordinating a national-scale validation network in China. The prototype of this network has been established; 12 anchor stations over seven representative regions were chosen as basic observatories. Over the past few years, various ground-based remote sensing experiments and intensive field campaigns were individually and/or jointly conducted at four core sites, in order to investigate remote sensing and validation mechanisms and theories. To facilitate the operational validation task, a web-based validation system (LAPVAS) is designed to use multi-scale in situ and remote sensing data for validating quantitative RSPs at a regional or global scale. This system is built with assembled modules and contains a validation database subsystem, an RSP accuracy evaluation subsystem, and an inter-external interface.

Research progress related to each component of the proposed VS is rewarding, as it improves our understanding of challenging issues in validation, especially in answering the question of how to acquire reliable ground truth at the pixel scale over heterogeneous land surfaces for validating RSPs. Future research should first further explore the validation methodologies, including validation metrics, the quantitative assessment of land heterogeneity, and uncertainty occurring during the validation process. Great importance could be attached to scaling approaches, which could facilitate ground truth acquisition at the pixel scale and provide the essential information required for remote sensing retrieval and validation. Second, more reference sites could be included in the validation network in order to improve the network to be more applicable for large-scale validation practice.
Finally, joint efforts involving the international community are anticipated to deliver international specifications for validating land RSPs, to better fulfil the requirements of the validation activities of global-scale RSPs.

Acknowledgments: This research was supported by the Natural Science Foundation of China projects (grant numbers: 91425303 and 41101320), the National Key project by MOST (grant number: 2016YFC0500100), and the inter-disciplinary Innovation Team of the Chinese Academy of Science. We would also like to thank three anonymous reviewers and the editors for their constructive comments and suggestions.

Author Contributions: Shuguo Wang, Xin Li, and Yong Ge conceived the structure of this paper. Shuguo Wang prepared the manuscript. Xin Li supervised the work. Yong Ge was responsible for the development of validation approaches. Rui Jin, Mingguo Ma, and Shaomin Liu conducted various validation activities and provided related materials, and were involved in coordinating the prototype of the national-scale validation network with Xin Li. Qinhuo Liu and Jianguang Wen designed and established the operational validation system. All authors approved the final manuscript.

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

References


105. Huang, G.H.; Li, X.; Ma, M.G.; Li, H.Y.; Huang, C.L. High resolution surface radiation products for studies of regional energy, hydrologic and ecological processes over Heihe River Basin, northwest China. *Agric. For. Meteorol.* 2016. [CrossRef]


111. Wang, X.F.; Ma, M.G.; Li, X. Validation of MODIS GPP product at ten flux sites in northern China. *Int. J. Remote Sens.* 2013, 34, 587–599. [CrossRef]


118. Parajka, J.; Bloschl, G. Validation of MODIS snow cover images over Austria. *Hydrol. Earth Syst. Sci.* 2006, 10, 679–689. [CrossRef]


166. Li, X.; Jin, R.; Liu, S.M.; Ge, Y.; Xiao, Q.; Liu, Q.H.; Ma, M.G.; Ran, Y.H. Upscaling researches in HiWATER: Progresses and prospects. *J. Remote Sens.* 2016. [CrossRef]

