vegetation index design, to build empirical models, machine learning techniques, or simply used as lookup tables for parameter retrieval.

7.2.1 Water Stress Detection

Water stress occurs when the demand of water exceeds the available supply during a certain period or when poor water quality restricts its use. It is well known that severe water deficits affect many physiological processes and have a strong impact on agricultural yield (Hsiao et al., 1976). However, even moderate water deficits, which are not easy to detect, can also have important negative effects on yield (Hsiao and Bradford, 1983). It is important to be able to assess the level of stress through pertinent indicators. The early detection of water stress is a key issue to avoid yield loss, which can be affected even by short-term water deficits (Hsiao et al., 1976).

Water stress affects plant spectral signal through three main ways: changes in the photosynthetic processes, pigments, and heat dissipation (e.g., xanthophyll cycle and chlorophyll fluorescence); direct changes in reflectance features associated with liquid water in leaf tissues, and changes in the canopy temperature consequence of a decrease of the evaporative cooling associated with stomata closure. Signal changes under stress conditions are located in specific spectral areas and in some cases do not surpasses the 3% of the signal under non-stress conditions (Figure 7.1, from Zarco-Tejada et al., 2000).

Up to date, radiative transfer models are not fully integrating these plant protection mechanisms. The processes and pigments affected by pre-visual water stress vary quickly throughout the day, therefore, complete collection of data describing the photoprotection processes is complex as all measurements need to be done almost concurrently (i.e., leaf spectra, stomatal conductance, water potential, photosynthetic rate, fluorescence, and leaf destructive sampling for pigment quantification). As the resources involved in field data collection are many, the benefit of sharing existing data is enormous.

The measurement of the fluorescence component of the plant functional apparatus is gaining increased attention with the development of the fluorescence explorer (FLEX) satellite mission (Vicent et al., 2016). An important prerequisite of the success of the mission is the establishment of

![Figure 7.1](image.png)

**Figure 7.1** Leaf reflectance of dark adapted leaves (dotted line, first measurement) and under steady-state condition (continuous line, last measurement), showing the reflectance difference associated with blue fluorescence, chlorophyll fluorescence, and xanthophylls pigment cycle. (From Zarco-Tejada, P. J. et al. 2000. Remote Sensing of Environment 74(3): 582–595.)
a spectral database to provide experimental evidence of the reliability of the atmospheric correction schemes developed for FLEX. Data from selected specialized point spectrometers (FloX; JB Hyperspectral Devices) mounted on towers (Figure 7.2) will be ingested into a spectral database and linked with HyPLANT (Rossini et al., 2015) and Sentinel-2 satellite data within the AtmoFLEX project in the framework of the ESA Earth Observation Envelope Programme EOEP-5.

7.2.2 Nutrient Stress and Pigment Content Detection

Indices for nutrient content are generally derived at the leaf, or patch scales in the case of pastures, based on the specific spectral absorption features (Thulin et al., 2014). The difficulty of diagnosing specific nutrient stress lies in the overlapping absorption spectral features for many primary nutrients (Suárez and Berni, 2012). Another difficulty associated with nutrient stress detection is the use of indices derived from leaf spectral data for canopy spectra. The structure of the canopy and the background characteristics play a very important role in the overall canopy signal (Suárez et al., 2008). For that reason, some indices derived at the leaf scale cannot be used at the canopy scale. Some authors apply index combinations to account for the background and structural effects (Haboudane et al., 2002), but this technique is not as straightforward for every case. Coupling leaf and canopy spectral databases with the pertinent metadata would bridge this gap and provide robust spectral indices for nutrient content detection at all scales.

The estimation of pigment content has been helped by spectral databases since radiative transfer models include chlorophyll concentration as input parameter (PROSPECT, Jacquemoud and Baret, 1990). Once the simulation of unlimited leaf spectra covering all ranges of input parameters was possible, the search for optimal vegetation indices could start. Lookup tables allowed the design of specific experiments and new generation sensors based on their capability to represent vegetation spectral responses (Richter et al., 2012).

Radiative transfer models have been inverted to derive parameters (Meroni et al., 2004 and others; Darvishzadeh et al., 2008; Jacquemoud et al., 2009) but this method can only be applied to existing model inputs. Figure 7.3 shows how lookup tables built from leaf-canopy coupled model
simulations can be used to derive algorithms to compute chlorophyll content from a vegetation index (from Berni et al., 2009a).

There are many pigments that are still not integrated in leaf models and are of interest for specific physiological processes. Another limitation of the use of radiative transfer modelling is the inability to simulate realistic scenes or constrain the model inputs properly avoiding ill-posed solutions (Zurita-Milla et al., 2015). Apart from general models derived from extensive lookup tables built from radiative transfer model simulations, locally derived empirical models are a common approach. The latter is very accurate within a small area but lacks of applicability in others. More generalised empirical models can be built integrating existing data collected over multiple experimental sites with inputs of all vegetation types if existing data collections were shared. If that was the case, the parameterisation of radiative transfer models with leaf optical properties could heavily benefit as well from rich data collections of spectral data with associated pigment concentrations to allow for a stratified trait allocation dependent on leaf position within the canopy.

### 7.2.3 Disease Detection

Disease detection is one of the most critical elements of plant management, especially for agricultural crops. Plants can be affected by abiotic stresses like shortage of water or nutrients; or by biotic stresses like insects, fungi, bacteria, or viruses. In many cases, the exact cause of the disease can only be discerned through destructive sampling, but knowing the symptoms in an environmental, spatial, and historical context can help pinpoint the cause.

Regional and federal organisations are fostering collaborations to keep spread of disease as limited as possible and avoid large-area infestations. The collaborative initiatives stem mostly from producers safeguarding the industrial sector and consumer safety. However, there is little knowledge interchange between the scientific community and agricultural crop growers. Hence, there is a great potential in data collected in affected fields that could help the early detection of affected areas. Examples of how spectral changes can lead to disease detection are shown in Figure 7.4.

Studies focusing on the detection of specific diseases in vineyards (Blanchfield et al., 2006; Renzullo et al., 2007; Meggio et al., 2010), fruit orchards (Calderón et al., 2013), grains (Mewes et al., 2011),
and herbaceous crops like cotton (Camargo and Smith, 2009) exist. The further application of the results of these studies is spatially limited (local studies) and in many cases the detection is sensor and species specific. The scientific community focuses on the results and apparently does not yet realise the potential of raw data collected across species and territories. Mohanty et al. (2016) presented a study where over 54,000 crowdsourced red-green-blue (RGB) pictures helped identify 26 diseases in 14 crop species. This study encouraged other authors and Wang et al. (2017) and Fuentes et al. (2017) repeated the technique for disease severity in apples and tomatoes, respectively, the following year. These results show how powerful databases can be for detecting crop damage. These techniques do not, however, allow early detection as they are based on visible symptoms and use limited spectral data. Studies have demonstrated that hyperspectral data can be used for early detection of diseases (Rumpf et al., 2010; Serranti et al., 2017) and, therefore, the assimilation of spectral data with disease-specific metadata in spectral databases could inform detection hypothesis testing and algorithm development.

Some diseases are systemic, affecting the whole plant system. In these cases, plants present similar symptoms to the ones of water stress, closing stomata and reducing evaporation and photosynthetic efficiency in the short term and vegetation growth and wilting in the long term (Beaumont, 1995). Other diseases present symptoms in the leaf surface in the form of spots or discolouration. In those cases, remote sensing can assist in detecting the pigmentation differences. Lastly, there are diseases that present both systemic effects and discolouration symptoms.

Since determining the cause of plant disease can be very difficult, spectral libraries of leaf and canopy measurements for different species and disease levels accompanied with complementary metadata may establish a good starting point for investigating the best remote detection techniques.
7.2.4 Food Security: Food Grading, Quality, and Source of Origin

In a time of mass agricultural production dominated by large corporations supported by political processes, food security is of global concern. Impacts of climate change, compounded by environmental shocks such as floods, and a decreasing amount of arable land, all serve to emphasise the need to ensure a more resilient agricultural production system (Turner et al., 2017). In Australia, the need to evaluate and monitor the impact of pests and diseases on agricultural productivity, seen to be of high risk to food biosecurity, has been identified as a high priority area to help manage risk (Craik et al., 2017).

The evaluation and assessment of food quality control and safety is a highly topical area of research related to food security and strategies which attempt to correct flaws in the global food system by making agricultural production more secure and sustainable for all. While well established, food chemistry studies are dominated by the use of FTIR microscopy centred on spectra of infrared absorption and emission. The development of spectral libraries across the full-range optical spectroscopy range (350–2500 nm) offers complementary and expanded absorption and reflectance data yielding information on water status, cellulose, and lignin for studies in food quality. Detailed spectra collected across a range of pure to contaminated agricultural crop products could contribute to consistent analytical methods to further our understanding of introduced contaminants and spoilage to food and food products. Grains and meals used for livestock feed are susceptible to contamination and spoilage, with potential to affect human consumption (Shen et al., 2016). Benefits would include the development of comprehensive detection screening methods to detect additives. Further, grain-based spectra, whether sub-products for biofuel production or livestock feed, can be used to determine food origins and traceability (Tena et al., 2015), an important component within strategies such as food sovereignty which emphasise food security and sustainability. Investigations which shift research into food grading, quality control, and origin traceability into comprehensive SIS accompanied by full-range optical spectral libraries is a clear area of potential research as humans enter into an increasingly complex global food system.

If spectral libraries are built from consistently acquired, full-range spectra with sufficient metadata to determine fitness for purpose, it has been shown that spectral matching algorithms, better known as successful for semi-automated mineral mapping, have potential to be effective for agricultural crop classification of hyperspectral imagery, particularly if data are collected to include seasonal and phenological state (Ndamanuri and Zhell, 2011). Teluguntla et al. (2017) provided an example of a carefully constructed field-based data collection which encompassed a range of crops and conditions over extensive areas. Class-based spectra could be generated using a quantitative spectral matching technique (QSMT) based on spectral correlation similarity to sort spectra into classes subsequently compared to an ideal spectrum for each class. While part of a satellite-based, spatially extensive automated cropland classification process, the conceptual framework and use of techniques rules could be a potential model for how to better address spectral matching across varying vegetation type and condition level. Rich ground-based spectra collection across crop types, condition, phenology state, and season could allow researchers improved use of machine learning algorithms and move semi-automated mapping of agricultural crop status to a new level.

7.2.5 Soil-Vegetation-Atmosphere-Transfer Models

Soil-Vegetation-Atmosphere-Transfer (SVAT) models compile the energy and water transfer processes on Earth into numerical equations based on the physical processes. These models are difficult to parameterise due to their complexity, especially when it comes to the vertical fluxes (Van Loon and Troch, 2001). Remote sensing data, as acquired systematically at the global scale, presents itself as convenient input to these models. Satellite information has been widely introduced as training and input data in the form of derived Leaf Area Index (LAI), or the fraction of Absorbed Photosynthetic Active Radiation (fAPAR), evapotranspiration, surface albedo, or temperature (Caselles et al., 1992; Olioso et al., 1999; de Wit et al., 2012).
SVAT models have been adapted to incorporate the whole energy balance and simulate resulting reflectance for all spectral regions (Wigneron et al., 1993; Olioso et al., 1999, 2002; Sobrino et al., 2011; Widlowski et al., 2011; Qiu et al., 2016). The full implementation of ecological models resolving the energy balance together with satellite sensor evolution puts us in the best position for global vegetation monitoring (Betts et al., 1996). There is still pending work in this sector and compiling the existing data would be of help. Other models like PROSPECT (Jacquemoud and Baret, 1990) were built upon spectral databases and have been widely used in direct and forward mode to inform vegetation assessment and monitoring. The inversion of radiative transfer models (RTMs) is very powerful once the simulation parameters have been properly defined. This step can be extremely complex for some vegetation types and the inversion is essentially ill-posed. This problem can be overcome by having spectral databases covering the full range of input trait parameters. In the past, databases based on model simulations have been used to derive vegetation indices (Haboudane et al., 2002). More recently, statistical models or emulators, based on databases, have been used to approximate the functioning of RTM inversions (Rivera et al., 2015).

Latest advances include the SCOPE model that incorporates the water and energy balance effects on vegetation physiology and the resulting signals including photoprotection mechanisms and resulting fluorescence signal (van der Tol et al., 2009). The performance of this model has been tested for many species and only failed to describe spectral variations related to xanthophyll pigment concentration (van der Tol et al., 2104). The next model generations would benefit from datasets covering all mechanisms and the model could be used to inform on global vegetation performance using fluorescence provided by the FLEX mission.

7.3 DATABASE IMPLEMENTATION AND SPECTROSCOPY DATA LIFE CYCLE

The actual implementation of a spectral database or spectral information system may vary to a certain degree, depending on the choice of software system components. We illustrate the common structure with the example of the SPECCHIO database (Figure 7.5).

The backend of the system is a relational database, storing spectral data and their metadata, as well as handling user accounts and user groups. Access to the database is strictly provided through a web service with a defined application programme interface (API) with most interfaces requiring a user authentication. Clients can access the SPECCHIO server by a variety of software tools, but always utilise the same API, written in Java. This allows communication with the server using either the SPECCHIO Java client or any programming language that provides a Java bridging technology. All lower level calls are handled by the Java components, presenting abstract data entry and access methods without requiring intrinsic knowledge of the database schema. This abstraction allows the running of optimisation routines to reduce metadata redundancy. Typically, spectroradiometer measurements are often replicated to establish the target variation. In these cases, most of the metaparameters are highly redundant and the system can reduce the stored metadata volume typically by around 30–60%, depending on the metaparameter set. This feature in turn increases the speed of data queries due to significantly reduced entries in the database.

Spectral information systems support the spectroscopy data life cycle to a large extent from: data acquisition where required metaparameters may already be pre-informed and aligned with available metadata attributes offered by the spectral information system; to loading, augmenting, and processing the data; to finally retrieving information to build and test new hypotheses, leading to potentially new experiments (Figure 7.6).

Metadata augmentation and data processing must be transparent to the user through the storage data provenance information. The system thus permits the storage of several different processing levels, for example, raw digital numbers, radiances, and reflectances, with clear dependencies and information about the processing algorithms. The handling of provenance adds to the replicability of scientific findings, constituting an important pillar of transparent science.