Gaussian Processes for Temporal and Spatial Pattern Analysis in the MISR Satellite Land-Surface Data

A dissertation submitted to the Faculty of Science,
University of the Witwatersrand, Johannesburg,
in fulfilment of the requirements for
the degree of Master of Science

Adrian John Cuthbertson

Supervisor: Professor Clint van Alten
School of Computer Science

Johannesburg, 30th May 2014

Abstract

The Multi-Angle Imaging SpectroRadiometer (MISR) is an Earth observation instrument operated by NASA on its Terra satellite. The instrument is unique in imaging the Earth's surface from nine cameras at different angles. An extended system MISR-HR, has been developed by the Joint Research Centre of the European Commission (JRC) and NASA, which derives many values describing the interaction between solar energy, the atmosphere and different surface characteristics. It also generates estimates of data at the native resolution of the instrument for 24 of the 36 camera bands for which on-board averaging has taken place prior to downloading of the data. MISR-HR data potentially yields high value information in agriculture, forestry, environmental studies, land management and other fields. The MISR-HR system and the data for the African continent have also been provided by NASA and the JRC to the South African National Space Agency (SANSA). Generally, satellite remote-sensing of the Earth's surface is characterised by irregularity in the time-series of data due to atmospheric, environmental and other effects. Time-series methods, in particular for vegetation phenology applications, exist for estimating missing data values, filling gaps and discerning periodic structure in the data. Recent evaluations of the methods established a sound set of requirements that such methods should satisfy. Existing methods mostly meet the requirements, but choice of method would largely depend on the analysis goals and on the nature of the underlying processes. An alternative method for time-series exists in Gaussian Processes, a long established statistical method, but not previously a common method for satellite remote-sensing time-series. This dissertation asserts that Gaussian Process regression could also meet the aforementioned set of time-series requirements, and further provide benefits of a consistent framework rooted in Bayesian statistical methods. To assess this assertion, a data case study has been conducted for data provided by SANSA for the Kruger National Park in South Africa. The requirements have been posed as research questions and answered in the affirmative by analysing twelve years of historical data for seven sites differing in vegetation types, in and bordering the Park. A further contribution is made in that the data study was conducted using Gaussian Process software which was developed specifically for this project in the modern open language Julia. This software will be released in due course as open source.

Declaration

I declare that this dissertation is my own unaided work, except where otherwise
acknowledged. It is being submitted to the University of the Witwatersrand, Johannesburg,
for the degree of Master of Science. It has not been submitted before for any degree
or examination in any other university.

Adrian John Cuthbertson
30th May 2014

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Multi-Angle Imaging SpectroRadiometer (MISR) is an Earth observation instrument operated by NASA on its Terra satellite [Diner *et al.* 2002]. This instrument is unique in that it comprises nine cameras, imaging the same locations on the Earth's surface from different angles distributed along the track of the satellite, providing multi-angled reflectance factors in electromagnetic bandwidths suitable for many applications.

A system known as MISR High Resolution or MISR-HR [Verstraete *et al.* 2012] has been developed by scientists at the Joint Research Centre of the European Commission (JRC) in Italy and implemented as a software package by Linda Hunt at the NASA Langley Atmospheric Science Data Centre in Virginia, USA. The purpose of the MISR-HR system is to derive many additional data values describing the interaction between solar energy, the atmosphere and the Earth's surface. It also generates estimates of data at the native resolution of the instrument (275 m) for 24 of the 36 camera bands for which on-board averaging (to 1.1 km) has taken place prior to downloading of the data. The MISR-HR data is useful in fields such as agriculture, forestry, environmental studies, land management and others.

The scale of the MISR-HR data and their value is highlighted by the following analogy: the MISR-HR system could also be viewed as a virtual collection of many thousands or even millions of independent instruments positioned above every 275 metre square area of the planet's surface, taking snapshot measurements and computing over 100 useful metrics for that square, from 3 to 10 times a month depending on the latitude.

During 2010, scientists from NASA, the JRC and the Council for Scientific and Industrial Research (CSIR) in South Africa, carried out field validations of MISR-HR in the Kruger National Park. With the success of this project, the MISR-HR system was made available by the JRC and NASA to the South African National Space Agency (SANSA), along with the MISR data for the whole of the African continent. The data constituted roughly weekly orbits from the year 2000 onwards. SANSA has installed the MISR-HR system and continues processing the data to generate the higher resolution datasets. These are provided in a format designed by NASA (HDFEOS), common to satellite data systems.

Satellite remote-sensing of the earth's surface is characterised by irregular sampling, outliers and missing observations in the time-series of data due to orbital constraints, instrumental artefacts, or atmospheric effects (e.g. cloud obscuration). This necessitates a process of smoothing the time-series and estimating missing data values, filling in gaps, discerning periodic structure and potential prediction beyond the boundaries of the time-series [Musial *et al.* 2011] and [Atkinson *et al.* 2012]. These studies reviewed and evaluated the most commonly used methods for addressing this problem, generally for the application of identifying phenology patterns in land-surface vegetation. The studies were motivated by bringing some consolidation across the many available time-series methods and in better informing researcher choices

of the most viable and accurate alternatives. Both studies concluded that the methods reviewed largely met the criteria that they were evaluated against, but all had particular advantages and disadvantages. The choice of method would largely depend on the analysis goals and on the nature of the underlying time-series structures.

In a different context, a long established statistical method known as Gaussian Processes (GPs) has recently become widely researched in the machine learning field [Rasmussen and Williams 2006]. This has also been applied to time-series analyses in a number of domains, but had not been identified in the above studies for satellite remote-sensing time-series methods. GPs had earlier been used in Geophysics spatial statistics work in a form known as Kriging [Cressie 1993], and other applications (for example, downscaling the pixel resolution in satellite images [Atkinson *et al.* 2008]). Gaussian Process Regression (GPR) has however also been used for time-series problems with a number of examples cited in [Rasmussen and Williams 2006]. Interesting case studies of GPR in time-series monitoring of tidal and weather sensors in coastal locations in the United Kingdom have been reported in [Roberts *et al.* 2013]. GPR was applied to optimal sensor positioning, sensor monitoring and providing rescue services with real-time information support.

The satellite remote-sensing time-series methods are generally rooted in a variety of mathematical techniques such as spline fitting, matrix algebra, signal processing and general statistical methods. These are typically of a form often classed as *parametric* models. Gaussian Processes are different, in being a form of stochastic processes. They are rooted more in probability theory and Bayesian statistical methods. These are of a class known as *non-parametric Bayesian* models.

With parametric methods the parameters for the constituent functions being modelled are usually specified or inferred as part of the modelling process. The functions either fulfill the purpose of relating the parameters to the data, or to an underlying process which generates the data. Gaussian Processes on the other hand involve placing probability distributions over the space of (possibly infinite) unknown functions which could explain the data, without specifying what those functions or their parameters are. GPs are also parameterised, but the parameters are rather used in controlling the probability distributions comprising the model. As such, they are rather referred to as *hyper-parameters* and the stochastic methods are termed non-parametric methods [Rasmussen and Williams 2006].

A key feature of GPs, is that best-choice hyper-parameters can be *learned* from the data and possibly from prior beliefs, assumptions or facts about the "shape" of the likely outputs of the underlying functions. This modelling process is also based on a *principled* probability framework and the same mathematical machinery can model a wide variety of applications. In particular, by forming the models from covariances of the probability distributions placed over the underlying functions, a level of abstraction is introduced and great variability in the models can be achieved within the single consistent framework.

The aims of the research presented in this dissertation have been to question whether Gaussian Processes can complement or add value to the above-mentioned satellite time-series methods. The research similarly therefore addresses the problem of irregularity and uncertainty in satellite time-series data, and also aims to further inform the choices available to researchers interested in these problems.

A subsidiary aim has been to adopt a research methodology which showcases Gaussian Processes as a powerful data analysis method and the MISR-HR system as a highly valuable source of land-surface information. A further aim, as part of the methodology process, has also been to develop a set of practical software components which can subsequently be used to perform actual time-series work in the MISR-HR data and other satellite remote-sensing systems.

The research questions for this study are based on the evaluation criteria adopted for the satellite method reviews in [Musial *et al.* 2011]. These included testing whether accurate estimates could be made for

missing or irregular observations in order to fill the gaps. Also whether noise in observation values could be adequately dealt with, and whether a smoothed differentiable curve could be fitted in order to re-sample the time-series for further analysis. The evaluations also considered the ability of the methods to discern structural characteristics of the time-series, such as whether underlying periodicities could be modelled for further applications such as identification of vegetation phenology events. Various methods for evaluating a measure of "goodness-of-fit" were also used in the evaluations. From these criteria the authors formed a cohesive set of requirements that such time-series methods generally, could or should fulfill. These time-series requirements have been considered in forming the research questions to be assessed in applying Gaussian Processes to time-series problems in satellite remote sensing data.

The research methodology comprises two approaches. Firstly to present the background to the currently used satellite remote time-series methods and a detailed exposition of the GPR method as it can be applied to time-series problems. Secondly thereafter, to show a data-based case study of the application of GPR to time-series analysis in the MISR-HR data obtained from SANSA for the Kruger National Park area in South Africa.

In order to provide a logical flow to the connections of the various topics with the research objectives, the dissertation has been laid out in the following manner. Chapter 2 provides the research background ordered by: background on the MISR satellite instrument and MISR-HR system; background on current satellite remote-sensing time-series methods; a detailed exposition of Gaussian Processes and the GPR methods; and finally the software developed and utilised to extract data from MISR-HR and perform GPR processing on that data. Chapter 3 then lays out the research questions and the thesis for the dissertation. The case study is presented over two chapters, firstly in Chapter 4 to perform a GPR time-series analysis of *one* case-study site and to specifically assess that in terms of the research questions from Chapter 3. Thereafter, in Chapter 5 the broader case study will be presented covering details of the sites chosen, data extraction procedures and a more general presentation of GPR applied to various sites within the Kruger National Park and how GPR can be used to analyse those. Chapter 6 summarises and draws final conclusions from this work.

In concluding this introduction, the contributions sought and hopefully achieved by this project are: to confirm that Gaussian Processes are a powerful and consistent method with Gaussian Process Regression being fully viable as a satellite remote-sensing method and having differentiators which would add value to the currently used methods; to showcase the MISR-HR system and the potential value of its data in an example over the Kruger National Park; to introduce the Julia Scientific and Technical programming language as a viable platform for Gaussian Process applications generally and satellite time-series applications in particular; and last but not least as a viable platform for data processing applications with data from the MISR-HR system.

Chapter 2

Research Background

This chapter provides background for the research problem defined in Chapter 3, the methodology adopted in Chapter 4 and the case studies in Chapter 5. The MISR satellite instrument and the MISR-HR system are first described to provide the context for the application chosen for assessing the research problem and the data case studies. Time-series methods commonly used in satellite remote sensing are described and a general set of requirements for such time-series methods is noted. Further studies for current methods of vegetation phenology event monitoring are also outlined. Gaussian Process Regression is described in detail with sections providing the theory of the methods and examples. The data processing for MISR-HR datasets and the Gaussian Process numeric computation have been done with software developed specifically for this project using the Julia language. The motivation for doing this and the software developed are described in some detail.

2.1   The MISR Satellite Instrument

The Multi-angle Imaging SpectroRadiometer (MISR) instrument operated by NASA on the Terra satellite is described in [Diner *et al.* 2002] and [Diner *et al.* 2007]. The Terra platform was launched on 18 December 1999 and the MISR instrument became operational on 24 February 2000. The technical information and data specifications are provided in [Jovanovic *et al.* 2012].

The satellite orbits in a polar sun-synchronous orbit which is precisely maintained with on-board propulsion. The orbit has a periodicity or *repeat period* of 16 days, in which the satellite repeats exactly the same orbit, observing the same locations from the same viewpoint. The time period required by each instrument on the Terra platform to cover the full surface of the Earth depends on the swath width of that instrument. For MISR, data need to be accumulated for 9 days to achieve complete coverage at the Equator. That period is much shorter for MODIS (about 3 days), as a result of that instrument having a wider swath.

The MISR instrument contains nine cameras, and observations are recorded for each location of the surface from nine angles spanning approximately $140°$. The camera identifiers and their respective angles are Df: $70.3°$, Cf: $60.2°$, Bf: $45.7°$, Af: $26.2°$, An: $0.1°$, Aa: $26.2°$, Bf: $45.7°$, Cf: $60.2°$, Df: $70.6°$. The suffix letters represent 'f' - forward, 'n' - nadir and 'a' - aft.

Detailed characteristics of the MISR instrument's recording features and data generation are described in [Diner *et al.* 1998]. MISR is known as a "push-broom" instrument and observations are recorded in four spectral bands, three for the visible bands and one for near-infrared (20-40 nm full-width at half height): blue (446.4 nm); green (557.5 nm); red (671.7 nm); and near-infrared (866.4 nm).

The approximate swath (view width) of the nine cameras along the orbit path is 385 km and the Ground Sampling Distance (GSD) along-track is 275 m for all cameras. Although the size of the projections of the sensor pixels on the ellipsoid varies greatly with camera angle, their effect is "removed" through a de-convolution step in the ground segment.

The raw datasets captured by the cameras are processed and re-sampled to the Space Oblique Mercator (SOM) projection. Although measurements from each camera are to 275 m resolution, the downloaded resolutions for the off-nadir cameras for blue, green and near-infrared are provided as averaged to 1.1 km resolution. This is known as the "Global Mode" of operation. All the nadir camera's bands and the red band for off-nadir cameras are however downloaded at 275 m resolution. The instrument can also operate in "Local Mode" where the full data for all cameras can be downloaded for specifically arranged areas. The MISR-HR system described below uses data acquired in Local Mode to verify a-posteriori, that the estimates generated by MISR-HR are in fact reasonable [Verstraete *et al.* 2012].

The data from the MISR instrument are processed by the NASA Langley Atmospheric Science Data Center and is made available publicly along with an assortment of derived data products. These support applications in many fields for both atmospheric and land-surface analyses.

MISR's ability to provide multi-angular imaging is unique to this satellite instrument and offers many advantages over nadir-only imaging systems. Some of these are described in the next section.

Diagram of MISR Operation

MISR Instrument
- 9 cameras
- nadir, front, rear angular views
- red, green, blue near-infrared images

Orbit Direction

Images Co-registered to surface points
- 36 (4 x 9) images - different angles
- red, green, blue near infrared images
  at 9 angles
- resolution 1.1 km (24/36), 275 m (12/36)
- local mode 275 m (36/36)
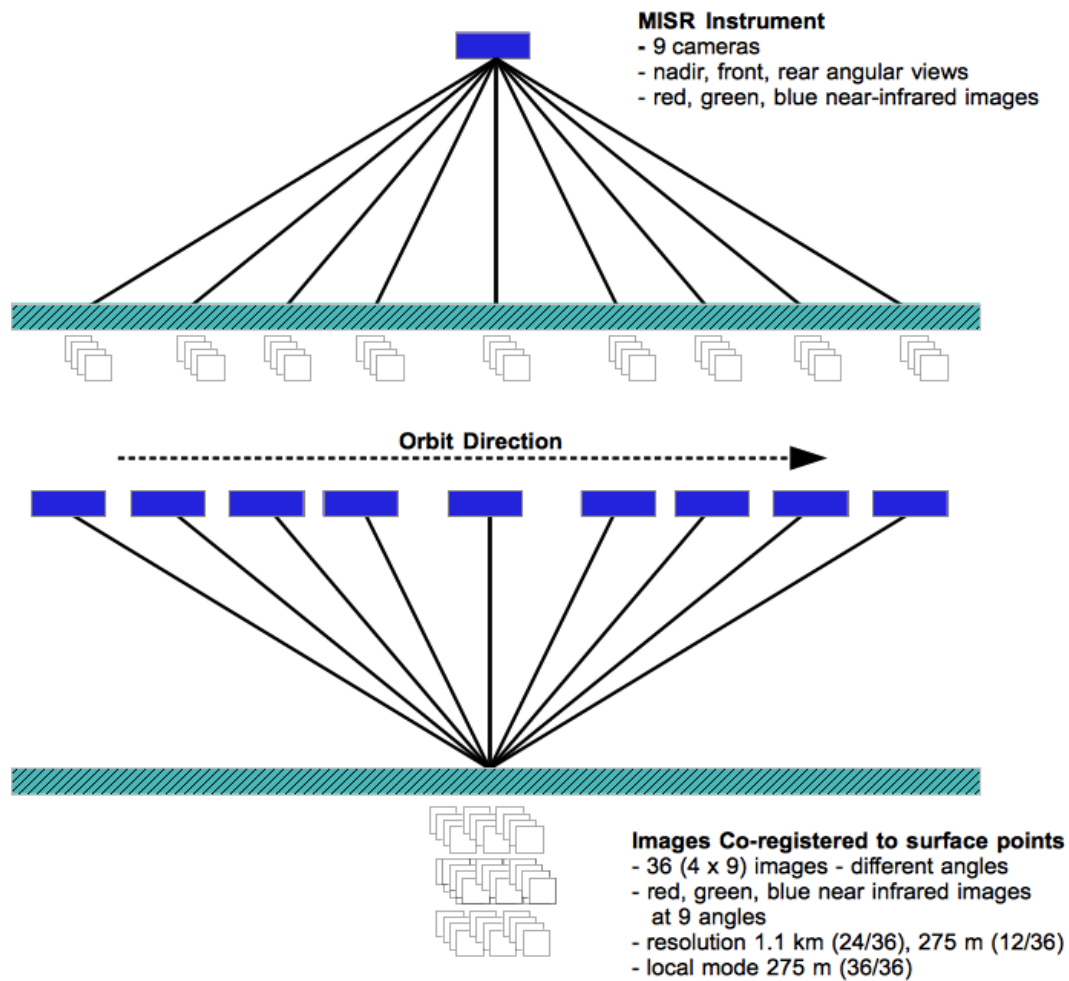
Figure 2.1: MISR Satellite Instrument Schematic.

The above figure shows a schematic of the MISR instrument and how the multi-angular images are processed to generate 36 spectro-directional values for the same location on Earth from the nine cameras[1].

2.2 The MISR-HR System

The MISR-HR system introduced in Chapter 1 is described in more detail in this section.

---

[1]Based on an original figure (Figure 4.2) in [Jovanovic *et al.* 2012].

2.2.1 Models and Products

MISR-HR generates extended data products from MISR, exploiting angular atmospheric reflectance factors from the standard data products, to estimate the averaged 1.1 km resolution to 275-metre pixels in the visible and near-infrared spectral bands. This effectively provides a 16-times increase in the resolution of the derived products over those available from the standard MISR set.

The data has also been further processed to generate many derived reflectance parameters and radiation flux metrics for each pixel. These have been shown in [Verstraete *et al.* 2012] to provide accurate ecological values which agree meaningfully with those produced by land-based remote-sensing and measurement alternatives.

This is of importance in studies of land-surface plant-biomass and surface change dynamics. The information is of value to researchers in conservation, forestry, agriculture, urbanisation, environmental studies and other fields.

The 2010 case study [Verstraete *et al.* 2012] shows favourable comparisons between the MISR-HR data and the data produced by instruments at the local CSIR Skukuza research station. Examples of raw time-series generated from the data demonstrate land-surface changes from possible causes such as encroachment of human settlements and anomalous events such as fires, land-clearing, drought damage and others.

Many satellite products for vegetation information are based on *vegetation-indices* which are empirical (as opposed to physically-based) relations which must be adjusted separately for each location and time-period, and cannot be generalised or applied anywhere else. By exploiting the multi-angular bidirectional reflectance factors (BRF), MISR-HR methods are designed to estimate the key vegetation parameters directly, in a rigorous way based on the geometries of the spectral reflectances between the sun, the surface and the satellite instrument. This, combined with the higher resolution images potentially provides more accurate vegetation information at a more detailed level than many of the other similar satellite systems.

The data products for MISR-HR are grouped in sets providing: top-of-atmosphere bidirectional reflectance factors (TOA-BRF); surface or bottom-of-atmosphere bidirectional reflectance factors (BOA-BRF); and values derived from two models known as the RPV and the JRC-TIP models. These models have been developed over many years and have a large body of literature describing their evolution. For purposes of the sample cases for this dissertation, we make use of specific products from the BOA-BRF, RPV and TIP datasets related to surface vegetation. The following descriptions are based on the high-level view of the products from [Verstraete *et al.* 2012], which in turn cites all the contributions made to the development of the models for MISR-HR.

The data values derived from the RPV model provide measures of the anisotropy of the surface BRFs, in the four spectral bands. The model parameters are k, a measure of the "bowl" or "bell" shape form of the anisotropy, ρ, a measure of the BRF amplitude and and Θ, a measure of the general tendency of the observed target to scatter light preferentially forward or backward. For k, a value of 1 means the surface is Lambertian. Values of k larger than 1 imply a bell-shaped anisotropy and k values lower than 1, a bowl-shaped anisotropy. In a field trip to Kansas to examine places with various pre-computed k values, it was confirmed that locations characterised by a bell-shaped anisotropy tended to correspond to areas where tall dark vegetation was scattered over relatively brighter ground, while the bowl-shaped anisotropy occurred over areas without such structural features. Although in [Verstraete *et al.* 2012] it is suggested that there are good logical and theoretical arguments to support the field study findings, this has not yet been widely confirmed with different field measurements in different ecosystems. Values of k therefore can be related to the shape of the anisotropy as a definite mathematical property, but perhaps to

vegetation heterogeneity or structure only with the qualification that there is just circumstantial evidence relating these.

Measures for *albedos* or bihemispherical reflectances (BHR) for broad and narrow visible and near-infrared spectral bands are computed from the RPV model. Although the term "albedo" is used by many researchers in different contexts with different meanings, in the MISR-HR context it refers to BHR values.

Finally, JRC-TIP (Joint Research Centre - Two-stream Inversion Package) generates a number of vegetation specific estimates for the radiation fluxes absorbed, scattered by the canopy and absorbed by the surface background. These are similarly provided in visible and near-infrared spectral bands. An effective leaf-area index (LAI) is also computed.

For RPV and the JRC-TIP parameter sets, the algorithms apply an inversion procedure which attempts to minimise the differences between model predictions and observations. A "cost" value of the minimization process is reported which gives some indication of the "goodness-of-fit" of the predictions. This may be an indication of outliers in the observations, but could also be caused by other reasons such as a lengthy exploration of the parameter search space. As will be seen in Chapter 4, there are simulated values which appear reasonable despite quite high cost values.

The MISR-HR parameters used for the sample cases for this dissertation are:

- RPV k in the red spectral band (also labeled as RPV-K in the sample cases), which is, subject to the above-mentioned qualification, a useful "characterisation" of the heterogeneity and diversity of surface vegetation in the spatial dimension, and also of seasonal and other changes in the temporal dimension.

- FAPAR is an estimate of the radiation absorbed in the visible spectrum - Fraction of Absorbed Photosynthetically Active Radiation. It is a direct estimate of the productivity of the vegetation canopy. Its time variations may be interpreted in terms of vegetation phenology.

- BHR-VIS and BHR-NIR - the albedos in the visible and near-infrared bands respectively. These measure the fraction of the incoming solar irradiance that is reflected back towards the zenith (sky) hemisphere by the surface. The values can be used directly in surface radiation or energy balance models. They do not however intrinsically contain any information about drought, erosion, fire or human activity such as bush clearing. In the sample cases, such interpretations are made about surface activity from observed temporal variations in albedo, but should be considered as conjecture in the absence of ancilliary data or proper validation studies, which have not been undertaken as part of this work.

- Costs as described and qualified above, for detecting possible inaccuracies and outliers in the data.

- BRFs in the blue, green and red spectral bands from the nadir camera for generating approximate visual spatial map images of the surface areas under study.

2.2.2 Datasets

The MISR-HR datasets are provided for blocks of pixels in the Space Oblique Mercator (SOM) projection from individual orbits of the satellite. The blocks follow the path of the orbit passing close to but not directly over the poles from north to south. The blocks are uniquely referenced by their Orbit Id, Orbit Date, Path Number and Block Number. Each block represents 512×2048 (275 m) "pixel" locations on the surface. There is some overlap in the longitudinally neighbouring blocks from neighbouring paths, although in latitude there is no overlap between successive blocks. The blocks are defined as areas extending beyond the actually observed swath for orbital and computational reasons. They do not stack

directly above and below each other: each one is slightly shifted (westward) by an amount that is latitude dependent. The dataset HDF files containing the block data also contain ancilliary information about this variable shift. Each pixel could be measured in up to three different orbit paths. However, pixels at the edges of the blocks far from the orbit swath are marked as "edge" pixels and are not usable.

Similarly, the MISR-HR algorithms are carefully designed to mark pixels as "obscured" and unusable if not all data from each camera is available. This could be due to measurements not being viable due to camera angles unable to view surface topography angles, or due to atmospheric or environmental obstacles such as clouds, smoke, pollution, etc.

2.3 Time-Series Methods in Satellite Remote-Sensing

As the aim of this dissertation is to explore the viability of Gaussian Processes as a method for temporal pattern analysis, in this section we review methods typically used in satellite remote-sensing applications for smoothing the time-series of acquired observations and other purposes such as gap-filling, outlier identification, interpolation and prediction, as well as discovering underlying causal processes or facilitating higher-level goals, such as phenology.

Recent evaluations of commonly used methods were undertaken by [Musial *et al.* 2011]. The methods evaluated were primarily for the application of time-series in analysis of phenology.

Satellite instruments now provide global observation information dating back over many years, or even decades. A common problem, is that observations are often compromised by geophysical processes such as cloud cover or other atmospheric conditions, limited sunlight in the polar regions in winter periods and by various surface environmental or human factors. The study was aimed at evaluating common methods in this problem area and informing researcher choices in the selection of suitable methods.

In addition to reviewing the methods, [Musial *et al.* 2011] developed a sound set of criteria for the functional requirements of suitable time-series methods and the problems they should address. These are suggested as the determining factors for an optimal underlying mathematical model which would capture essential properties of the system in a systematic way. The properties of such a model could be physical or statistical.

We first outline the methods reviewed by [Musial *et al.* 2011], after which their evaluation criteria are examined in some detail as these are important to the research questions established for this dissertation in Chapter 3. The methods are also described in some detail in order to illustrate the types of models used and their parameterisation, learning or optimisation approaches and means of determining accuracy of the method.

2.3.1 The Smoothing Spline Method

This method [Hutchinson and de Hoog 1985], [Reinsch 1967], [Whittaker 1923], allows the formulation of a continuous curve between the points of a time-series by joining cubic polynomials at calculated "knot" points, which allow the first and second derivatives to be continuous throughout. A single parameter λ dictates how closely the curve fits the observation points. As $\lambda \to 0$, the curve passes through all the points and thus reduces to basic interpolation. As λ increases, the curve becomes "smoother" until as $\lambda \to +\infty$, it becomes a linear least-squares fit. An optimization method for determining an optimal value for λ, has been applied [Craven and Wahba 1979], where Generalized Cross Validation (GCV) is used to iteratively remove points and statistically determine the best λ value [Hutchinson 1988]. The

GCV value is formulated from a weighted sum of the discrepancies between the removed point and a curve fitted to the other points.

The smoothing spline method is considered attractive in that it is a simple and efficient method which applies a function directly to the points themselves and makes no assumptions about the structure of the time-series or the causes of the variations by the underlying processes generating the observations.

2.3.2   The Singular Spectrum Analysis (SSA) Method

This method, proposed by [Kondrashov and Ghil 2006] is aimed at formulating the structure of a time-series as a composition of more elementary components representing various sub-features. The method as described and evaluated in [Musial *et al.* 2011], involves a sequence of transformations of a multidimensional "trajectory" matrix comprising vectors representing partial views of the original time-series. These are iteratively processed through a set of varying length windows of the series and an assortment of matrix manipulations are performed in order to structure the original time-series as a superposition of a trend, some harmonic oscillations and noise. A final averaging procedure produces Hankel matrices which become the trajectory matrices of the underlying time-series. The original time-series can be reconstructed and used to regenerate missing values and fill the gaps. The method is applicable to multivariate time-series, but only univariate series were considered in the [Musial *et al.* 2011] evaluations.

The SSA method has parameters m representing a window length factor and η representing a threshold related to a number of orthogonal functions used in the matrix manipulations. A cross-validation method is used to find suitable values for the parameters by iteration over alternatives after removing portions of the data and determining the best statistical (RMSE) values for the parameters.

SSA is considered suitable for filling gaps where the time-series have non-harmonic oscillation shapes. It has advantages over more traditional methods based on classical Fourier spectral analysis, which may need numerous alternative phase and amplitude trigonometrical functions to provide a valid result. SSA however, may be constrained by the high computational requirements due to the necessary matrix manipulations.

2.3.3   The Lomb-Scargle Method

This method developed in [Hocke and Kämpfer 2009], [Press *et al.* 1992], [Lomb 1976], and reviewed by [Musial *et al.* 2011], involves removing the mean of the overall original time-series from each observation and enhancing the spectral information by applying a Hamming window [Harris 1978]. A periodogram known as the Lomb-Scargle periodogram is derived producing the equivalent of fitting a least squares of sine and cosine functions to the original observations [Hocke and Kämpfer 2009]. The significant components, based on a parameterized threshold, are used to reconstruct a signal to which a reverse Hamming window is applied, finally resulting in a complete and continuous signal. Missing values can be determined and gaps filled by resampling at the desired frequency. In some applications of the method accuracy may be reduced near the boundaries of the series, but this may be alleviated by the alternative use of a Kaiser-Bessel window [Harris 1978]. In this case, a further "shape" parameter is required. The significance threshold parameter is either simply set as a fixed fraction of the calculated maximum value in the periodogram, or done by means of a confidence level analysis. The Lomb-Scargle method is considered suitable for time-series with strong periodic structure, although the observations do not need to be evenly distributed in time.

2.3.4 Criteria for Choices of Time-Series Methods

In [Musial *et al.* 2011], a time-series is defined as a finite set of ordered pairs of numerical expressions, each pair comprising a reference to a point in time and a value for an observation or measurement at that time. The distinction is drawn between continuous time-series for data collected by analog instruments and those in which discrete records have been either digitised from analog streams or recorded at specific times. These are also referred to as discrete time-series. The sources of satellite remote-sensing observations for vegetation phenology are typically discrete records and the methods evaluated are thus for discrete time-series. Two broad requirements are identified in both establishing the evaluation criteria and in the choice of the above methods for evaluation.

The first requirement is that estimates of likely values of a time-series should be able to be made at arbitrary times in order to fill gaps from missing data or irregularly acquired observations. This falls within the general problem of interpolation or "curve fitting", where a model forms a curve through or near the observed points and can be used to calculate estimates at other arbitrary points. In order to make useful estimates, the curve should be reasonably "smooth", with some "flexibility" in its closeness to the observed points. There should be some compromise between the smoothness of the curve and the closeness to observed points. This compromise should have some measure, typically known as a "goodness-of-fit" measure, which can be used to inform choices for any model parameters controlling that compromise. The differentiability of the model function should also be known or knowable and the time-series able to be re-sampled at frequencies suitable for the application.

Methods meeting this broad requirement do not necessarily make any assumptions about the underlying mechanisms or processes actually causing the variability in the time-series. The advantage of this is that such methods could be applied to models of arbitrary complexity and could give satisfactory results independently of any changes in time to the underlying processes. The disadvantage is that the dynamics of the underlying processes and the properties causing variability are not necessarily "learned" within the model, possibly constraining the forecasting abilities of the time-series.

The second broad requirement (depending on the application), is that the method *should* allow for the modelling of the underlying processes in some manner, to discern periodicities *and* aperiodicities, in order to make reasonable forecasts in time, beyond the observed records.

2.3.5 Comparing Goodness-of-Fit

In addition to establishing the requirements for time-series methods, [Musial *et al.* 2011] also considered the importance of the method of comparison of the goodness-of-fit measures when comparing different models. The RMSE (root mean square error) and RMSD (root mean square deviation) statistics over the model are often used, but the MAE (mean absolute error) statistic was considered superior for the evaluations in this study. Although the statistics each give a measure of the difference between modelled values and observations, MAE had been shown to better assess the overall fit of the model, while the others over-emphasised large individual differences.

In [Musial *et al.* 2011], the methods were evaluated against the above broad requirements for both artificially generated sample data and a number of actual time-series datasets for vegetation FAPAR (Fraction of Absorbed Photosynthetically Active Radiation), sunspot observations, atmospheric carbon-dioxide concentration records and the Dow Jones Index. The datasets were also chosen for their particular representation of some of the problems informing the above requirements.

2.3.6 Other Methods Recently Studied

Other studies have evaluated different time-series to those mentioned above and also methods related directly to the determination of vegetation phenology.

Phenology generally, is the study of recurrences in natural phenomena [Verstraete *et al.* 2007]. In the context of plants or vegetation it is the study of those events reflecting growth and cyclical or seasonal development phases. This study reported the development of a detailed method for automatically determining key vegetation phenology events.

For space-based monitoring of vegetation phenology the goals are to monitor general phases of variability for the time periods where vegetation maintains particular types of activity, such as photosynthesis or the redistribution of plant material.

Important to systematic monitoring of these activities is the need to cater for different applications having different criteria for determining phenology events. In agriculture for example, crop growth seasons are related to periods in which air temperatures are above certain specified levels, or perhaps the period between the last frost (of a given severity) of the previous season and the first of the next season. Another example might be in the domain of CO_2, for monitoring the periods where net absorption of carbon is occurring in the vegetation.

[Verstraete *et al.* 2007] point out the usefulness of separating the identification of key events using suitable statistical methods from their relationship to the underlying requirements of an application. Their goal was to develop a procedure for the robust automatic statistical detection of key phenology events, using remote-sensing data for a number of vegetation types and ecosystems. The method involved analysis of decadal (ten-day) time-series in three phases of: determining the overall statistical properties of the time-series; fitting an S-shaped mathematical model through a moving window across the series; and analysing the results to determine key phenology events.

Another recent related study is reported in [Atkinson *et al.* 2012]. Their aims were also to evaluate a number of time-series smoothing methods, in this case for satellite remote-sensing data over the Indian sub-continent for vegetation phenology analysis. The methods were Fourier Analysis, Asymmetric Gaussian, Double Logistic and the Whittaker Smoother. Each method was tested against multiple goodness-of-fit statistics. These included RMSE (root mean square error), RE (residual error), AIC (the Akaike Information Criterion) and BIC (the Bayesian Information Criterion). Their evaluations also concluded that the methods were largely successful, but with different advantages and disadvantages dependent on the requirements.

2.3.7 Conclusions

The preceding sub-sections have illustrated that there are a good number of choices for time-series smoothing methods for satellite remote-sensing, for assessing and comparing their effectiveness and also for determining phenology events. The general conclusion of the evaluations cited was that the methods each have particular advantages and disadvantages. The choice of a suitable method largely depends on the specific goals of the application or research and on the properties of the underlying time-series.

The purpose of this dissertation is to consider whether given these many available satisfactory alternatives, Gaussian Process methods offer some differentiation justifying adding them to the already many choices available.

In particular in this section, we have drawn on the general requirements for time-series methods set out by [Musial *et al.* 2011] for (but not limited to) satellite remote sensing applications. These form the basis of the research objectives for the dissertation to be set out in Chapter 3.

2.4 Gaussian Processes

A key motivation for the research behind this dissertation has been the exploration of Gaussian Processes (GPs), a powerful Bayesian statistics method, reviewed in the canonical work of [Rasmussen and Williams 2006]. It has a long history in statistics and geostatistics (where it is known as "kriging"), and more recently in machine learning for applications in Bayesian regression and classification problems. There are also strong ties between GPs and other machine learning methods such as feed-forward neural networks, support vector machines, principal component analysis and others. In many cases GPs can be viewed as a generalisation or abstraction of those methods [Rasmussen and Williams 2006], [Bishop 2006], [MacKay 2003].

GPs are a class of *stochastic processes* in which some underlying phenomena or functions are "wrapped" in random variables (probability distributions), which collectively are a finite subset of a greater infinite probability distribution. GPs exploit the fact that when the infinite distribution is multivariate Gaussian, inferences can be made by only considering the subset and not the full infinite joint distribution. This probabilistic approach provides for modelling many types of underlying functions in non-linear and multi-dimensional spaces.

Depending on the size of the concrete subset distribution, inferencing methods can be computationally expensive, but with advances in both computation processor technologies and algorithm efficiencies, computational demands of GPs are no longer constraints for many problems and applications.

As the core of this dissertation describes the application of GP regression for time-series modelling in satellite remote sensing and for the MISR-HR system in particular, we provide some detail on the workings of GPs for this context. In the following subsections, we briefly review what machine learning is and then how the non-parametric, stochastic GPs fit within that space.

2.4.1 Supervised Machine-Learning Parametric Methods

Generally in supervised machine-learning, the goal is to exploit mappings from inputs (typically empirical data) to known outputs in order to make inferences or predictions about mappings from other inputs with unknown outputs [Alpaydm 2010].

In many supervised learning approaches, the methods attempt to fit the mapping to some function by finding suitable parameters to the function. This may be represented by $Y = f(X|\theta)$ where X are input values (each possibly a vector of values), Y are the known output values and $f(X|\theta)$ is the function of X given a set of parameters θ resulting in Y, or possibly $Y = f(X|\theta) + \eta$ where η is an element of randomness representing noise in the observations.

These are known as *parametric* methods. When the output Y is continuous, the model is known as *regression* and when it is discrete, *classification*. The machine-learning algorithms then optimize the parameters given a training set $D = \{(x_i, y_i)|i = 1, ..., n\}$ in order to achieve the best fit or correct classification for the function. Finally, inference or prediction algorithms implement $Y_* = f(X_*|\theta_*)$ where Y_* is the predicted value, X_* the specified input with unknown output and θ_* the trained parameters.

2.4.2 Gaussian Process - A Bayesian Non-Parametric Method

An alternative is the *Bayesian non-parametric* approach where the function generating the mapping from input data to observed outputs is considered to be unknown [Rasmussen and Williams 2006], [Bishop 2006], [MacKay 2003]. A probability distribution over the space of possible functions is formed in a Bayesian stochastic process. This comprises an infinite set of unknown random variables as a GP prior distribution, conditioned by an indexed subset representing observations or evidence. Bayes' theorem is applied to infer a posterior distribution, which is also a GP, from which predictions can be made.

A multivariate Gaussian distribution is constructed to contain a set of Gaussian random variables representing the *probabilities* that some unknown function has values Y at inputs X and possibly a noise factor representing uncertainty or noise in the observations:

$$p(\text{Y}) = \mathcal{N}(\mathbf{M}, \mathbf{C}). \tag{2.1}$$

The distribution is represented by an $n \times n$ covariance matrix $\mathbf{C}$ and a mean vector $\mathbf{M}$ of length n, where n is the number of elements in X. The observation noise factor is typically incorporated into the covariance matrix diagonals which represent the variances of the individual variables. A GP then, provides for constructing this distribution.

Following the definitions in [Rasmussen and Williams 2006][2], a GP is a collection of random variables, any finite subset of which has a joint multivariate Gaussian distribution. It is fully specified by a mean function and a covariance-kernel function defined as:

$$\mathbf{m}(\text{X}) = \mathbb{E}[f(\text{X})], \tag{2.2}$$
$$\mathbf{c}(\text{X}, \text{X}') = \mathbb{E}[(f(\text{X}) - \mathbf{m}(\text{X}))(f(\text{X}') - \mathbf{m}(\text{X}'))], \tag{2.3}$$

and

$$f(\text{X}) \sim GP(\mathbf{m}(\text{X}), \mathbf{c}(\text{X}, \text{X}')). \tag{2.4}$$

Generally, X may be a set of scalar values over which f could be applied, or an *index set* of the possible values, such as a vector in $\mathbb{R}^D$. Often for time-series, X is simply continuous in $\mathbb{R}$ representing points in time directly.

Creating a GP model for a set of observations or training pairs $(\mathbf{x}, \mathbf{y})$, involves applying the mean and covariance functions and generating a multivariate Gaussian from those.

The properties of a multivariate Gaussian which facilitate the Bayesian non-parametric regression approach are: the *joint distribution* of any non-empty subset of its random variable components also forms a Gaussian; the *marginal distribution* of any subset is also Gaussian; and the *conditional density* of a subset, given the values of another subset, is also Gaussian. Importantly, these properties facilitate performing inference analytically and the posterior distribution can be obtained directly through computations based on matrix algebra, rather than (necessarily) having to resort to approximate inferencing methods.

---

[2]We adopt the notation $\mathbf{c}$ for the covariance kernel function and $\mathbf{C}$ for the covariance matrix, rather than the often-used $\mathbf{k}$ and $\mathbf{K}$, to avoid confusion with the RPV $\mathbf{K}$ notation used in the MISR-HR system.

The following shows how the posterior distribution is formed using these properties, to add the appropriately computed row and column vectors and sub-matrix to complete the joint distribution, which includes the prediction/s;

$$
\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{x}) \\ \mathbf{m}(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{c}(\mathbf{x},\mathbf{x}) + \sigma_n^2 I & \mathbf{c}(\mathbf{x},\mathbf{x}_*) \\ \mathbf{c}(\mathbf{x}_*,\mathbf{x}) & \mathbf{c}(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix} \right), \tag{2.5}
$$

where $\mathbf{m}$ and $\mathbf{c}$ are the mean and covariance functions defined previously, $\mathbf{y}$ are the observation values (including noise), $\mathbf{x}$ are the observation locations, $\mathbf{f}_*$ are the predictive means, $\mathbf{x}_*$ are the prediction locations, $\sigma_n^2 I$ is the (Gaussian) signal noise or likelihood uncertainty (often added as a model hyper-parameter) with I as the identity matrix.

The following two equations can then be derived from this augmented joint distribution [Rasmussen and Williams 2006]. Here for brevity, we use $\mathbf{C} = c(\mathbf{x},\mathbf{x})$, $\mathbf{C}_{**} = \mathbf{c}(\mathbf{x}_*,\mathbf{x}_*)$, $\mathbf{C}_* = \mathbf{c}(\mathbf{x},\mathbf{x}_*)$ and $\mathbf{C}_*^T = \mathbf{c}(\mathbf{x}_*,\mathbf{x})$

$$
\mathbf{f}_* = \mathbf{C}_*^T(\mathbf{C} + \sigma_n^2 I)^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{x})), \tag{2.6}
$$

$$
\mathbb{V}_* = \mathbf{C}_{**} - \mathbf{C}_*^T(\mathbf{C} + \sigma_n^2 I)^{-1}\mathbf{C}_*. \tag{2.7}
$$

where $\mathbf{f}_*$ is the prediction mean and $\mathbb{V}_*$ is the prediction variance.

It should be noted that for most applications the mean function is not exploited and typically a mean-zero function is used which simply returns a vector of zeros. This is also the case for all the GP models formulated in this dissertation.

2.4.3 Goodness-of-Fit - The Negative Log Likelihood

In order to obtain a measure of the "goodness-of-fit" for the predictions, the *marginal likelihood* can be determined. This is the marginal likelihood of the observations, given by the integral of the likelihood times the prior [Rasmussen and Williams 2006]. The integration can be performed analytically, with the log marginal likelihood being computed from the results of the previous section by:

$$
\log p(\mathrm{Y}|\mathrm{X}, \theta) = -\frac{1}{2}\mathrm{Y}^T(\mathbf{C} + \sigma_n^2 I)^{-1}\mathrm{Y} - \frac{1}{2}\log|\mathbf{C} + \sigma_n^2 I| - \frac{n}{2}\log 2\pi. \tag{2.8}
$$

In practice (and further within this dissertation), the *negative* log likelihood (NLL) is computed rather than the log likelihood, which emphasizes the closeness or "minimisation" of the fit:

$$
NLL = \frac{1}{2}\mathrm{Y}^T(\mathbf{C} + \sigma_n^2 I)^{-1}\mathrm{Y} + \frac{1}{2}\log|\mathbf{C} + \sigma_n^2 I| + \frac{n}{2}\log 2\pi. \tag{2.9}
$$

The NLL then, is a measure of how good the predictions are, given the prior covariances subject to the values of hyper-parameters and the evidence of the training data. The lower (more negative) the NLL, the better the fit and the higher the likelihood of the posterior predictions.

The NLL can also be used in optimising (minimising) the hyper-parameters, which in GPs is the process of "learning" the hyper-parameters from the model and the data, using equation 2.9 as the optimisation objective funtion.

More detail on learning methods is provided in a later section.

In order to give context to the sections which follow, a simple example of a GP regression model is shown in the next section.

2.4.4 GP Regression Example

To show the workings of a GP model we assume a very simple time-series, in which there are 4 observations over a period scaled (say) from 0.0 to 2.0. The scale of the observation values is also 0.0 to 2.0. The observations are:

$$\mathbf{x} := [0.3; 0.6; 1.5; 1.6], \mathbf{y} := [0.2; 0.4; 0.9; 0.7] \tag{2.10}$$

We will also assume that after training with these input observations, we wish to predict the underlying function values at two points:

$$\mathbf{x}_* := [0.8; 1.2] \tag{2.11}$$

We will use a Squared Exponential covariance kernel function

$$c_{SE}(x_i, x_j) = \sigma_v^2 exp(-\frac{|x_i - x_j|^2}{2l^2}), \tag{2.12}$$

which takes two hyper-parameters, l, a "length" factor controlling the decay of the influence (covariance) of points near each other (LS), and σ_v^2, a "height" factor controlling the signal variance (SV) of the fitted predictions:

$$l := 0.5, \sigma_v^2 := 0.8 \tag{2.13}$$

Equations 2.6 and 2.7 also utilise a "signal noise" hyper-parameter σ_n^2 which is the uncertainty or (Gausian) noise associated with the observations. This is also termed the likelihood uncertainty (Lik) which is used in the plots to distinguish this from covariance hyper-parameters:

$$\sigma_n^2 := 0.001 \tag{2.14}$$

For this example (and generally in this dissertation), we do not make use of the mean function and thus:

$$\mathbf{m}(\mathbf{x}) := \mathbf{0} \tag{2.15}$$

Using the Squared Exponential covariance function, the computations can now be performed for equations 2.6 and 2.7:

$$\mathbf{C} + \sigma_n^2 I = \begin{bmatrix} 0.6401 & 0.5346 & 0.0359 & 0.0218 \\ 0.5346 & 0.6401 & 0.1267 & 0.0866 \\ 0.0359 & 0.1267 & 0.6401 & 0.6273 \\ 0.0218 & 0.0866 & 0.6273 & 0.6401 \end{bmatrix} \tag{2.16}$$

$$\mathbf{C}_* = \begin{bmatrix} 0.3882 & 0.1267 \\ 0.5908 & 0.3115 \\ 0.2402 & 0.5346 \\ 0.1779 & 0.4647 \end{bmatrix} \tag{2.17}$$

$$\mathbf{C}_{**} = \begin{bmatrix} 0.6400 & 0.4647 \\ 0.4647 & 0.6400 \end{bmatrix} \tag{2.18}$$

and finally the prediction means $\mathbf{f}_*$ and variances $\mathbb{V}_*$ are computed from equations 2.6 and 2.7:

$$\mathbf{f}_* = \begin{bmatrix} 0.67152 \\ 1.14371 \end{bmatrix} \tag{2.19}$$

$$\mathbf{V}_* = \begin{bmatrix} 0.0144 \\ 0.0200 \end{bmatrix} \tag{2.20}$$

and the prediction variances including the signal noise or likelihood uncertainty σ_n^2:

$$\mathbf{V}_* + \sigma_n^2 = \begin{bmatrix} 0.0145 \\ 0.0201 \end{bmatrix} \tag{2.21}$$
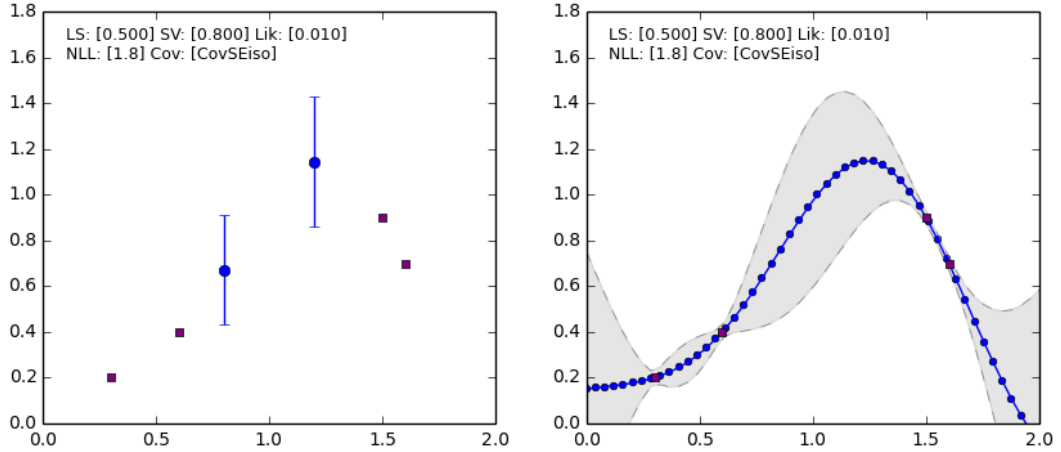
Figure 2.2: Sample regression model - Squared Exponential covariance function.

The left panel in Figure 2.2 shows the plot for the above computations with the observations as purple squares and the prediction points shown in blue with error bars indicating 2 standard deviations $2(\mathbf{V}_* + \sigma_n^2)^{0.5}$. The right panel shows the plot for the same model, but this time for 50 predictions evenly spread across the time-scale. The points are joined in the plot to form a curve, and the variance (also 2 standard deviations) is shown as a gray band[3].

The plots clearly show how the uncertainty reduces to the observation noise near the observation points, but diverges to the prior uncertainty specified by the covariance kernel function away from the observations. This is a good example of a Bayesian model at work - prior beliefs, conditioned by evidence

---

[3]Generally further in this dissertation, we will compute a "high density" of prediction points and then only show the predictive curve and the variance bands.

or observations, giving posterior predictions, but with a systematic measure of the posterior uncertainty.

The computations from the above example are based on algorithm 2.1 page 19, from [Rasmussen and Williams 2006] and are implemented here in Julia (to be described later). For completeness in this section, we provide a listing of the above computations in Julia source code (which line-for-line is almost identical to the algorithm):

```
x=[0.3;0.6;1.5;1.6]
y=[0.2;0.4;0.9;0.7]
xs=[0.8;1.2]

covfn=CovSEiso(0.5,0.8)         # instantiate the covariance function
sn2=0.01^2                      # the model likelihood (signal noise)
mu=fill!(similar(x),0.0)        # won't use the mean function

C=doCov(covfn,x,x)              # compute C
L=chol(C+(eye(n)*sn2),:L)       # obtain Cholesky decomposition (C==L'L)
alpha=L'\(L\y)                  # Cholesky factorisation

Cs=doCov(covfn,x,xs)            # C_star
fs=Cs'*alpha                    # f_star -> predictive means
v=L\Cs                          # solve for v
Css=doCov(covfn,xs,[])          # C_star_star
Vfs=Css-v'*v                    # V_star -> predictive variances
Vmdl=Vfs+sn2                    # With signal noise (Lik)
```

The procedure for using Cholesky decomposition instead of matrix inversion (equations 2.6 and 2.7) is more performant and provides better numerical stability. It should also be noted that the inversion of $(\mathbf{C} + \sigma_n^2 I)^{-1}$ is only dependent on the kernel computation for the training points. We can thus perform the inference process just once for a training set, and then perform different prediction sets (re-sampling) without repeating the training inference matrix inversion. This was done in generating the right-hand plot in Figure 2.2.

2.4.5 Covariance Kernels

[Rasmussen and Williams 2006] point out that in supervised learning, there is the crucial notion that for training records where a known target value represents some function of an input value, there is an assumption that other nearby input values are likely to have similar target values. With GPs this likelihood is represented in a systematic way by the covariance kernel function. The kernel determines the covariances between any two constituent variables of the overall distribution, or equivalently, the kernel can be viewed as a function of any two sample variables, say x_i and x_j.

If the "distance" between x_i and x_j is informative in explaining the difference between the observation values at those points, this can be modelled by *stationary* kernels which are dependent on $x_i - x_j$ (where x_i and x_j could be vectors in $\mathbb{R}^D$). When the kernel is a function of $r = |x_i - x_j|$, this models an *isotropic* influence of the distance between the points. When the kernel is only a function of r, these are known as *radial basis functions* (RBF). There are also *anisotropic* forms of stationary kernels, and various forms of *non-stationary* kernels.

A multivariate Gaussian has the property of its covariance matrix being a Gram matrix which is symmetric positive semidefinite (PSD). The only requirement of a kernel function therefore is that it generates a valid symmetric PSD matrix. Such kernels are referred to as PSD kernels. The properties of PSD kernels allow them to be combined as sums and products of other valid kernels: the sum of two kernels is a PSD kernel and the product of two kernels is a PSD kernel.

The functions for covariance kernels usually have "free" parameters controlling their properties. However, for Gaussian Processes they (and other model parameters) are termed *hyper-parameters* to clarify

that they are the non-parametric parameters of the stochastic model.

Covariance kernel functions allow much flexibility and many kernels have been developed to suit different modelling needs [Rasmussen and Williams 2006], [Abrahamsen 1997], [Duvenard *et al.* 2013].

We next outline some kernels which are suitable for time-series problems [Roberts *et al.* 2013], [Rasmussen and Williams 2006]. Although there are many kernels described in those works, the discussion here will be limited to just those kernels chosen for purposes of this dissertation. For purposes of reference for later sections and chapters, each kernel described is referred to by an identifying name. The kernels to be used are named CovSEiso, CovMaterniso, CovPeriodic, CovSum and CovProduct.

In the following subsections:

$$r = |x_i - x_j| \tag{2.22}$$

Squared Exponential Kernel

The commonly used Squared Exponential (CovSEiso) kernel is represented by:

$$c_{SE}(x_i, x_j) = \sigma_v^2 exp(-\frac{r^2}{2l^2}) \tag{2.23}$$

where the hyper-parameters are: l which is the *effective length scale* (LS) and σ_v^2 which is termed the *signal variance* (SV).

The following figures are examples of random samples drawn from a GP prior distribution (no training inputs), with the CovSEiso covariance function and the hyper-parameters (LS and SV) set as indicated above each panel. The samples are each actually for 100 **x** axis values evenly spread between -3.0 and $+3.0$. The **y** axis represents the variance, also ranging from -3.0 to $+3.0$. The GP has a zero mean-function. The blue curve is a continuous plot line connecting the points.
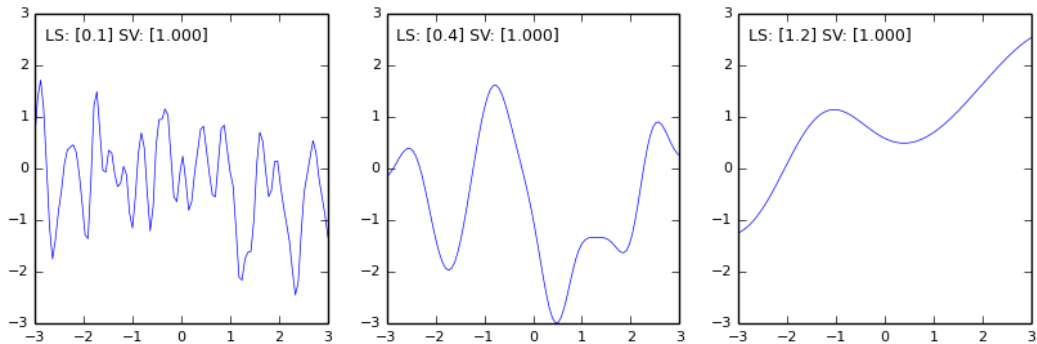

Figure 2.3: GP Prior Samples Drawn from a Squared Exponential Kernel - Example 1.
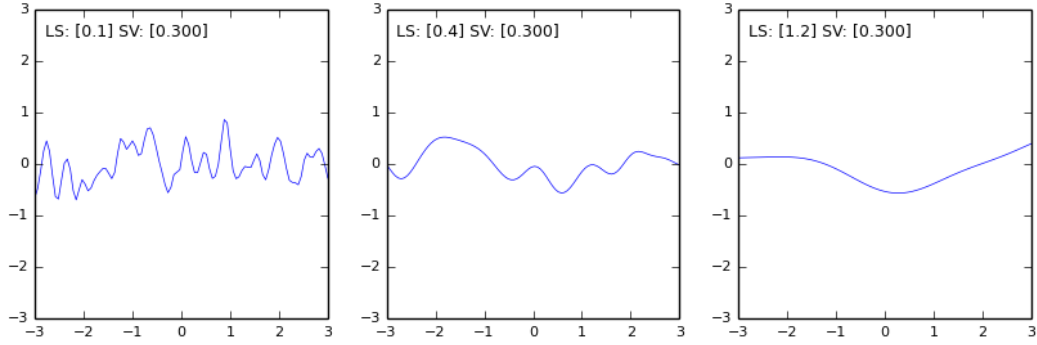
Figure 2.4: GP Prior Samples Drawn from a Squared Exponential Kernel - Example 2.

As can be seen from these plots, the prior covariances are controlled by the hyper-parameters. The CovSEiso kernel is a good general purpose kernel and is useful for a first attempt to get a sense of the structure in the data. It is however, often too "smooth" for some applications.

Matérn Kernel

The Matérn kernel has more flexibility in specifying the "roughness" of the covariances. It is represented by:

$$c_{Mat}(x_i, x_j) = \sigma_v^2 \left(\frac{1}{\Gamma(\boldsymbol{V}) 2^{\boldsymbol{V}-1}} \right) \left(2\sqrt{\boldsymbol{V}} \frac{r}{l} \right) \mathbb{B}_{\boldsymbol{V}} \left(2\sqrt{\boldsymbol{V}} \frac{r}{l} \right) \tag{2.24}$$

where the hyper-parameters are also l, the *effective length scale* (LS) and σ_v^2, the *signal variance* (SV). $\Gamma()$ is the standard Gamma function and $\mathbb{B}()$ a modified Bessel function [Rasmussen and Williams 2006]. $\boldsymbol{V}$ is a *class* parameter which when constrained to half-integer values $\boldsymbol{V} = 1/2, 3/2, 5/2, 7/2$, provides kernels with increasing degrees of differentiability. Our implementation treats these as different classes of Matérn covariance functions by instantiation with a suitable constructor parameter.

Generally, $\boldsymbol{V} = 3/2$ is utilised as it is well suited for rougher models than the SE kernel.

Figure 2.5: GP Prior Samples Drawn from a Matérn Kernel - Example 1.
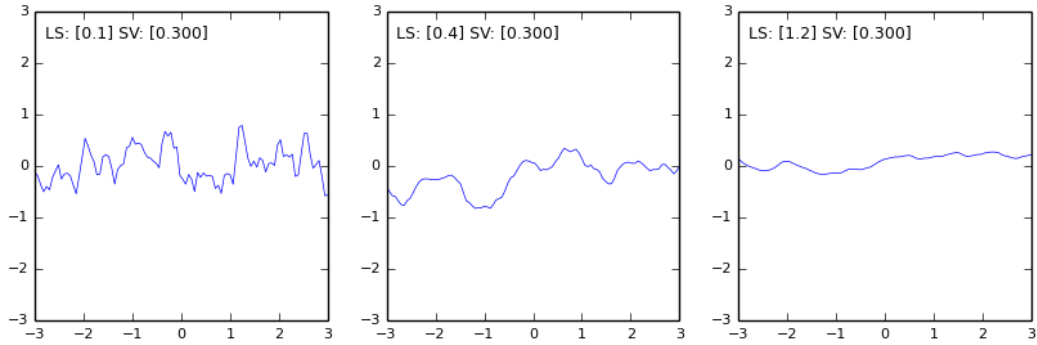
Figure 2.6: GP Prior Samples Drawn from a Matérn Kernel - Example 2.

Figures 2.5 and 2.6 show the rougher covariance effects.

Periodic Kernel

A general Periodic kernel (CovPeriodic) provides for periodicity in the covariance influences between points. It is represented by:

$$c_{Per}(x_i, x_j) = \sigma_v^2 exp\left(-\frac{2}{l^2}sin^2\left(\frac{\pi r}{p}\right)\right) \tag{2.25}$$

where the hyper-parameters are l, the *effective length scale* (LS), σ_v^2, the *signal variance* (SV) and p is the period (P). The length scale in this case is relative to the period and for larger values of l, the effects become sinusoidal, while for small values of l, the effects are of more complex harmonics.
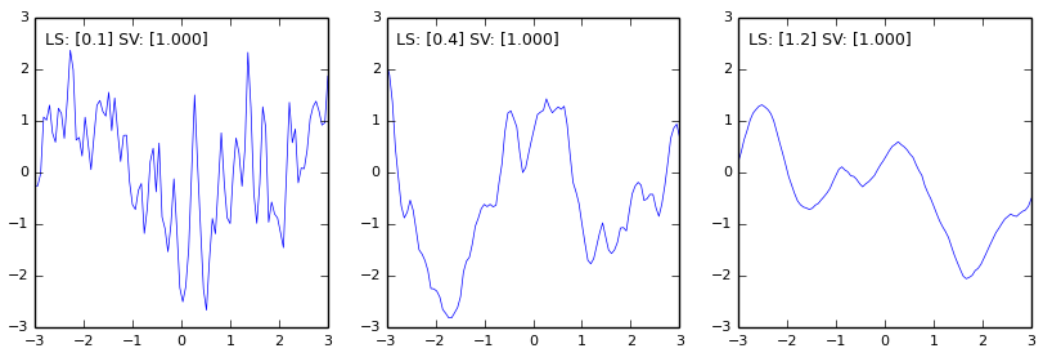
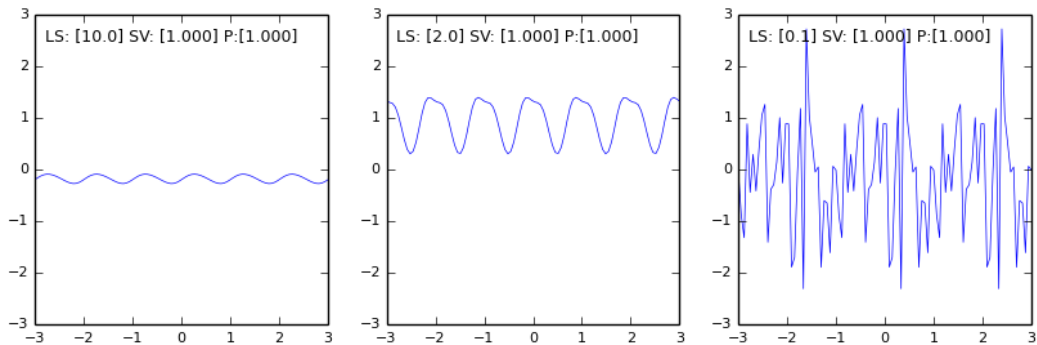

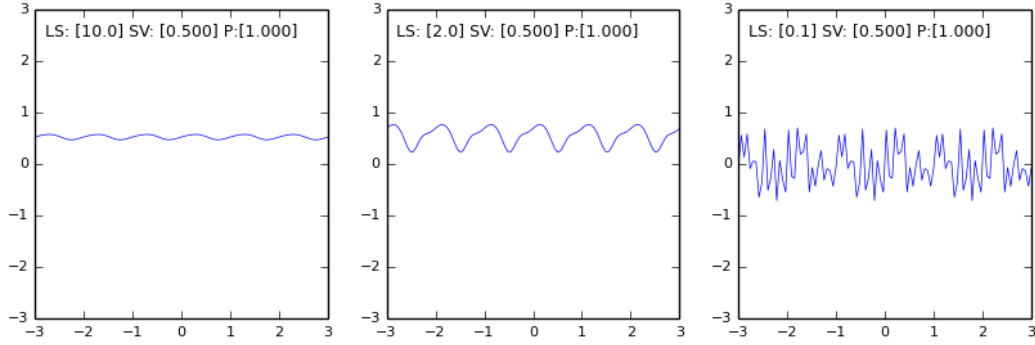Figure 2.7: GP Prior Samples Drawn from a Periodic Kernel - Example 1.

Figure 2.8: GP Prior Samples Drawn from a Periodic Kernel - Example 2.

Figures 2.7 and 2.8 show the periodic covariance effects. In each case the period is 1.0 with the varying length scale showing the sinusoidal and more complex harmonics effects. The signal variance affects the amplitude.

Composite Kernels

Composite kernels can be formed by summing two kernels, or taking the product of two kernels, and in each case the resultant kernel is positive semi-definite (PSD) and is thus a valid covariance kernel. We can also then form more complex kernels from combinations of products and sums of kernels.

For this dissertation, in Chapters 4 and 5 we sum the Matérn and Periodic kernels in order to discern reasonably complex periodic structure in the MISR-HR data.

Generally, when using covariance kernels, it is often difficult to choose appropriate hyper-parameters and the process of hyper-parameter optimisation or "learning" is required.

2.4.6 GP Hyper-Parameter Learning

In section 2.4.2, equation 2.9 for the negative log likelihood (NLL) was used for measuring the likelihood of the predictions. This can also be used for optimisation of the hyper-parameters for a GP model. This is the basic procedure for learning in Gaussian Processes, although much more sophisticated full Bayesian learning methods are provided in [Rasmussen and Williams 2006], along with various alternative approaches. There is also a substantial body of other research work dealing with this topic.

For purposes of this dissertation optimisation is based on the method of determining the hyper-parameters that minimise the NLL, with equation 2.9 as the objective function.

Minimisation is performed using a non-linear optimisation library, in this case NLopt [Johnson]. The COBYLA (Constrained Optimization BY Linear Approximations) algorithm from the NLopt library is used for derivative-free optimisation [Powell 1994] and [Powell 1998].

The (Julia) covariance kernel functions developed or ported for this project have also catered for computing derivatives for hyper-parameters for the CovSEiso, CovMatern and CovPeriodic kernels. A gradient-descent algorithm Low-storage BFGS (L-BFGS), also provided by the NLopt package has also been utilised [Nocedal 1980] and [Liu and Nocedal 1989]. The results of the two optimisation algorithms are similar, with the gradient-descent method being more computationally efficient. The derivative-free method however, is more flexible in handling composite kernels without the complexity of determining derivatives for multiple combinations of kernels.

2.4.7 Gaussian Process Algorithms Toolkit

A comprehensive toolkit, developed in Matlab/Octave, with many of the algorithms described in [Rasmussen and Williams 2006] and [Nickisch and Rasmussen 2008], is described in [Rasmussen and Nickisch 2010] and is available through their companion web site [The Gaussian Process Web Site].

This toolkit was invaluable in learning about Gaussian Processes and for testing various functions. However, it was decided that the Julia language would be used for implementing the GP software components used for this dissertation project.

2.5 Software Components and Libraries

The software components developed for this dissertation are described in this section. These include software functions for MISR-HR data extracts and processing, for Gaussian Process model processing and for plotting.

Although there are many software libraries and tools available which could have been used for many of the data processing and computational requirements for this project, it was decided to primarily use the more recently released (March 2012) Julia Scientific and Technical programming platform [Bezanson *et al.* 2012a].

The key reasons were:

- Gaussian Processes have not previously been implemented in Julia (as at the time of completing this dissertation), and a contribution could thus be made by doing so

- Julia has not been used for data and computational processing for the MISR-HR system, or as far as we know for satellite remote sensing applications generally, and a contribution could also be made in this arena

- Julia has a permissive Open Source license and any software developed for this project could be open-sourced and benefit wider communities

- Julia has an appealing proposition regarding its performance, ease-of-use for scientific applications and its ability to integrate well with other libraries

2.5.1 The Julia Scientific and Technical Programming Platform

The Julia language has been designed [Bezanson *et al.* 2012a] as a programming platform specifically for scientific and technical computing. It is positioned as a high-performance language, fully compiled with similar performance to C++, but with a high-level syntax and the ease of use of similar languages such as R, Python, Matlab and Octave. It also has many of the existing supporting libraries needed for scientific programming and many new ones exploiting modern computing capabilities such as concurrency across processor cores and computing clusters. The Julia web site is at [Bezanson *et al.* 2012b].

In [Lubin and Dunning 2013], Julia was used for implementing a new Algebraic Modelling Language (AML) for use in operations research and optimisation. Benchmarks were carried out for dense and sparse matrix computations used in their algorithms, comparing Julia with implementations in C++, MatLab, PyPy and Python. For these particular benchmarks, Julia's timed performance was within a factor of 2 of C++, compared with between 4 and 18 of C++ for PyPy and MatLab. Standard Python was 70 times that of C++.

2.5.2   MISR-HR Data Processing - EOS.jl Component

This component comprises a set of Julia functions wrapping selected functions from the HDFEOS library [The HDFEOS Libraries Web Site], originally developed by NASA and later made available as open-source. These functions are used for data extracts from the MISR-HR datasets and transformation of data to formats suitable for processing by the Julia numeric methods libraries. Also for conversions between the global longitude/latitude, SOM (Space Oblique Mercator) and the Block-Line-Sample coordinate systems used in MISR-HR.

MISR-HR Data Extracts and Transformation

Julia functions wrapping the data extract subset of the HDFEOS API (application programming interface) are: gdGridnames, gdOpen, gdAttach, gdInqdims, gdGridinfo, gdBlkSOMoffset, gdInqfields, gdProjinfo, gdFieldinfo, gdReadField, gdReadAttr, gdGetFillValue, gdDetach and gdClose. These are the essential functions required for accessing, interrogating and extracting data from all of the datasets available with the MISR-HR system.

Additionally, a number of Julia components provide support for transforming the MISR-HR row-major data array format to Julia (and the numeric methods libraries) column-major array format.

Various project-related extract utilities have also been developed for organising the data extracts for further computational processing: extractHistRPV, extractHistRPVSigmas, extractHistTIP, extractBlock, extractBlockL1B3, extractBlockRPV.

In terms of asserting correctness of the data extracts and format conversions, a utility known as HdfView is provided with the the HDFEOS libraries. This was used to view the original datasets and to verify that the data for various sample block coordinates, matched correctly with those generated by the Julia extractions and transformations.

Geolocation and Coordinate System Mapping

The NASA provided libraries also provide C-language templates for performing translations between various geospatial projection and coordinate systems. Based on these templates, Julia components have also been developed for the coordinate conversions required for the MISR-HR SOM, Block-Line-Sample and the Global Latitude and Longitude systems.

The functions implemented include: Misrproj, bls_latlon, coordQry, lonlat2som, som2lonlat, r2deg, dms2r, deg2r, dms2deg, deg2dms, bls2deg and deg2bls. These cater for degree-minute-second, decimal-degrees and radians.

Histogram for Outlier Filtering

As described earlier, certain MISR-HR records contain Cost fields which are a measure of computation costs of determining record values. High relative cost values can be an indication of possible outliers [Verstraete *et al.* 2012][4].

---

[4] The suggestion of using a histogram to filter outliers based on the Cost field was made by Dr Michel Verstraete of SANSA in a personal communication in regarding details of the MISR-HR data.

A standard Julia statistical histogram function was used in a component, OutlierFilter.jl to filter outliers for the MISR-HR datasets based on parameters for the field, the cost and the histogram options. This is utilised in Chapter 4 for filtering outliers for the RPV-K samples.

2.5.3 Gaussian Process Implementation - The GP.jl Component

The Gaussian Process Regression functions have also been developed in Julia, but with significant portions based on porting the required functions from the Gaussian Development Toolkit referred to previously.

Careful procedures were followed in replicating test sets from the Toolkit and ensuring that the Julia components produced identical results (within acceptable numeric tolerances). This was done by first porting a pseudo random number generator (gpml_randn.m) from the toolkit, specifically designed to be easily portable and useful for comparing identically generated random sample test sets across platforms. The Julia version (gpml_rand.jl) was then used to generate identical test data as the test cases provided with the toolkit, and the algorithms for the two platforms were then directly comparable.

A number of covariance kernel functions have been implemented as individual Julia components which can be "plugged into" the Julia GP model. These are shown in the following table.

Table 2.1: Covariance Kernels

Covariance Kernel	Component
Isotropic Squared Exponential	CovSEiso
Isotropic Matérn	CovMaterniso
Isotropic Rational Quadratic	CovRQiso
Periodic	CovPer
Linear	CovLin
Kernel Summing	CovSum
Kernel Products	CovProd

GP learning and optimisation is performed using the NLopt non-linear optimisation library as described in section 2.4.6. The author of the NLopt library (Steven G. Johnson) has also released a Julia package wrapping all of the available NLopt algorithms. These were used directly for the optimisation functions for this project.

2.5.4 Results Plotting - Julia and Python Matplotlib Libraries

Although Julia includes a number of plotting modules, it also contains a sophisticated interface to the Python system and to the Matplotlib plotting library [Hunter 2007]. These Julia packages were also developed and released by Steven G. Johnson.

A component plot.jl has been developed for this dissertation's plotting, leveraging the above-mentioned packages.

2.5.5 Time-Scale Model

A GPR model is most conveniently structured to work with continuous rather than discrete models. As MISR-HR data comprises discrete observations of the surface at intervals of days over periods of years,

a suitable time-scale model is required which combines this discrete observation time-scale with the continuous time-scale preferred for continuous GP algorithm computations.

This has been implemented as a support utility, using a standard Julia calendar library and functions to map between day-of-year integers and a 64-bit floating point value between 0.0 and 1.0. Providing a base year as a parameter, the utility gives a convenient and accurate method for mapping between a continuous numeric scale and a date-based scale over years or decades.

This is best illustrated by the Julia source code for the utility:

```
#### Utility tscal.jl ####
import Calendar

# Take a base year, a date string and return an x value
# relative to the start of base year, with a measure of 1.0
# being the x time for 1 year
#
function dt2x(baseyr,datestr)
        d=Calendar.parse("yyyyMMdd",datestr)
        x_yr = year(d)-baseyr
        enddoy = Calendar.dayofyear(ymd(year(d),12,31))
        return x_yr+float64((dayofyear(d)-1) / enddoy)
end

# Take a base year, and an x value and return the
# date string for the date relative to the x distance
# from the start of the base year
#
function x2dt(baseyr,x)
        x_yr=floor(x)
        yr=int(baseyr+x_yr)
        enddoy=Calendar.dayofyear(ymd(yr,12,31))
        doy=int((x-x_yr)*enddoy)+1
        Calendar.format("yyyyMMdd",ymd(int(yr),1,1)+days(doy-1))
end

# Create a type for holding the base year
type TSCal
    baseyr
end
# and functions to operate on that type
dt2x(tscal::TSCal,datestr) = dt2x(baseyr,datestr)
x2dt(tscal::TSCal,x) = x2dt(baseyr,x)

#---- Examples ---->
ts = TSCal(2000);

x1=dt2x(ts,"20100622")
==> 10.471232876712328

x2=dt2x(ts,"20100623")
==> 10.473972602739726

dx = x2-x1
==> 0.0027397260273982482

x3=x1+30*dx
==> 10.553424657534276

x2dt(ts,x3)              x2dt(ts,x3-11.0)
==> "20100722"          ==> "19990722"
```

The listing shows the functions defined first, then a Julia Type structure to hold a base year value,

followed by the functions to operate on that type (an example of Julia's multiple dispatch system). The imported Calendar is a general purpose Julia package for working with calendars.

After the function definitions, the examples show typical mappings between the real, continuous X-scale and the calendar (days) scale.

In later chapters, the dates for MISR-HR records from the orbits from 2000 through to mid-2012 are converted to an X-scale with a base year of 2000, which is then the X-scale for all GP processing. Predictions (and resampling) can be easily done for specified dates, by a simple mapping *external* to the GP algorithms which operate in the real-number space.

2.6 Conclusions

The background covered in this chapter has shown that the MISR-HR system is a rich source of sophisticated satellite remote-sensing data, which characterises all of the problems that time-series related studies might encounter. Current common time-series and vegetation phenology analysis methods have been reviewed and the general requirements for these methods established. Gaussian Processes have been covered in detail and their features and capabilities described, along with simple examples which give the context for the more sophisticated analyses to be done in Chapters 4 and 5. The Julia software platform and the software components developed for this project have also been described.

Chapter 3

Research Problem, Thesis and Methodology

3.1   The Research Problem

We have shown in Chapter 2 that satellite remote-sensing of the earth's surface is characterised by irregularity and uncertainty in the data due to atmospheric, environmental and other effects. To discern useful vegetation phenology information from the data, time-series of the data need to be enhanced to enable estimates for missing data values, reasonable interpolations made for gaps in the data, periodic structure discerned from underlying process causes and predictions made by extrapolating beyond the boundaries of the time-series. In each case, noise or uncertainty in the observations needs to be catered for.

3.2   The Thesis

The following premises drawn from Chapter 2 in the context of this research problem are:

- Previous work has identified problems characterising satellite remote sensing temporal data and defined requirements and solutions for time-series methods for temporal data analysis in these systems

- Recently evaluated commonly used methods all have advantages and disadvantages and choice of method is largely determined by the goals of the analyses

- The common methods are generally in the so-called parametric class of methods, often requiring different approaches to parameter determination, optimisation and accuracy validation

- Previous work has also shown Gaussian Processes to be well suited to time-series problems with similar characteristics to those required for satellite remote sensing

- Gaussian Processes, in the so-called non-parametric stochastic class of methods, have an advantage of being a principled framework based in probability theory, with the flexibility of different parametric time-series methods, but with a consistent approach to inference, prediction, specification and optimisation of hyper-parameters and accuracy validation.

We suggest based upon these premises, that Gaussian Processes can be applied to time-series smoothing in satellite remote sensing data for vegetation phenology analyses and further that all the problem requirements can be satisfactorily met within this single framework, additionally providing the advantages of consistency in hyper-parameter selection, optimisation and accuracy validation.

3.3 Research Methodology

A further premise drawn from Chapter 2 is that the MISR-HR data pertaining to vegetation phenology is characterised by all of the identified time-series problems typical of satellite remote sensing systems. In validating this thesis we will adopt the methodology of applying Gaussian Processes to sample cases from the MISR-HR data and assessing their effectiveness in addressing the time-series smoothing problems.

The Gaussian Process methods necessary for the sample cases have been implemented in the Julia scientific programming language platform as described in Chapter 2. Correctness of the implemented software components has been validated against the established GPML Gaussian Process toolkit described in Chapter 2. Data processing components for extracting and processing MISR-HR data from the available datasets have also been implemented in Julia, interfacing to the NASA provided HDFEOS libraries described in Chapter 2. Correctness of the data extraction and transformation components have been validated against an established viewer tool also available with the HDFEOS libraries. Plotting facilities from the Python Matplotlib library have been integrated with the Julia components for generating visual images of the MISR-HR data and plots of the Gaussian Process results.

An area of the Kruger National Park has been chosen for a data case study of the application of the Gaussian Process time-series methods to the MISR-HR data for that area. Sites have been selected from known vegetation landscapes and human encroachment border areas to facilitate meaningful comparative analyses in the Gaussian Process models for the sites.

One of the sites is specifically modelled to assess the models against the research questions. The sites are also comparatively modelled in order to further broadly support the thesis that Gaussian Processes are a viable temporal pattern analysis method and would be a worthy addition to the current common methods for vegetation studies in satellite remote sensing.

3.4 Research Questions

The principle research questions needing to then be addressed by this methodology are:

- Can Gaussian Processes be used to satisfactorily model time-series smoothing problems in MISR-HR data?

- Can the different time-series problems all be modelled within the single framework encompassing noise, prediction, hyper-parameter selection, model optimisation and accuracy validation?

More specifically, in modelling the MISR-HR data for sites in the Kruger National Park:

- Can estimates be made for missing data values?

- Can gaps in the data to be filled with reasonable interpolations?

- Can periodic structure in the underlying processes be discerned?

- Can plausible predictions be made by extrapolating beyond the time-series boundaries?

- Can noise and uncertainty in the data accuracy be catered for?

The next chapter provides an assessment of these questions.

Chapter 4

Gaussian Process Time-Series on MISR-HR Data

In this chapter we develop a number of Gaussian Process models for one set of MISR-HR data from the case study presented in Chapter 5. The goal here will be to test the GP methods on real data, but with the focus on assessing their abilities in terms of the research questions set out in Chapter 3. The case study in Chapter 5 will focus on showing the application of the methods to some explorations of vegetation phenology patterns.

As described in Chapter 2, RPV k is an important variable in the determination of other derived variables (in the TIP model). Also, subject to the subtleties and qualification of its interpretation as outlined in Chapter 2, in the red spectral band it gives a measure of the homogeneity of the surface. Higher values could indicate a rougher, more heterogeneous surface, while lower values could indicate a smoother, more homogeneous surface. In these models and in Chapter 5, we will only be considering k values for the red band. The variable in the plots and text will be labeled or referred to as RPV-K.

The data used in this chapter are for the RPV-K for site E1 from Chapter 5.

The first objective will be to generate a smoothed time-series from the data after which we will explore periodic structure and the ability to forecast values beyond the boundaries of the time-series.

4.1 Smoothing Time-Series

This section focuses on creating a smoothed time-series for the data which can be re-sampled to predict values for missing observations and fill gaps between observations.

4.1.1 Raw Time-Series

We begin by generating a plot showing the full history of the RPV-K data observations for the target site.
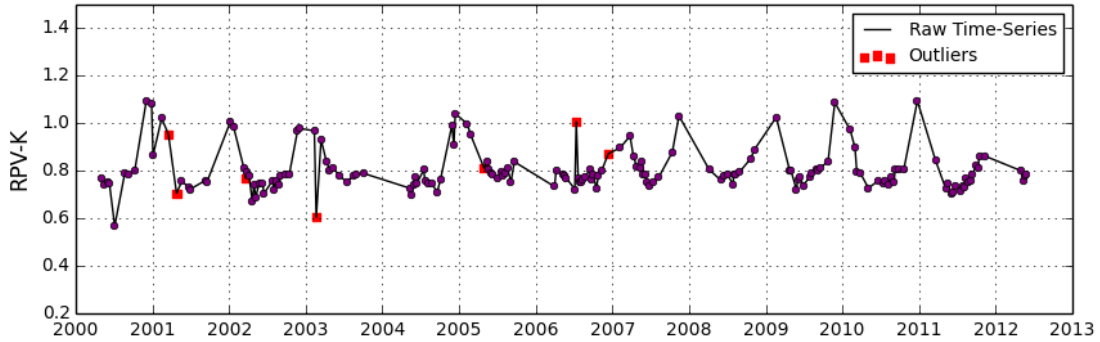
Figure 4.1: Raw Time Series for RPV-K with Outliers.

Figure 4.1 shows the the RPV-K observations over the time-scale from the year 2000 to mid-2012. For this chapter we use the full time-scale to have a broad view of the data while the models are formulated. More detailed views over shorter time-scales are shown in the case study in Chapter 5.

The cost field associated with RPV-K is a measure of the computational cost determining the predictions from the observations and could be an indication of outliers. We can use a standard statistical Histogram (see Chapter 2 for details) to statistically identify RPV-K values with high cost values, shown as red squares in the plot. These are then filtered and excluded from subsequent processing[1].

4.1.2 CovSEiso Kernel - Default Hyper-Parameters

The first smoothing process will be to run a Gaussian Process Regression (GPR) with the commonly used Squared Exponential covariance kernel (CovSEiso) with default hyper-parameters.
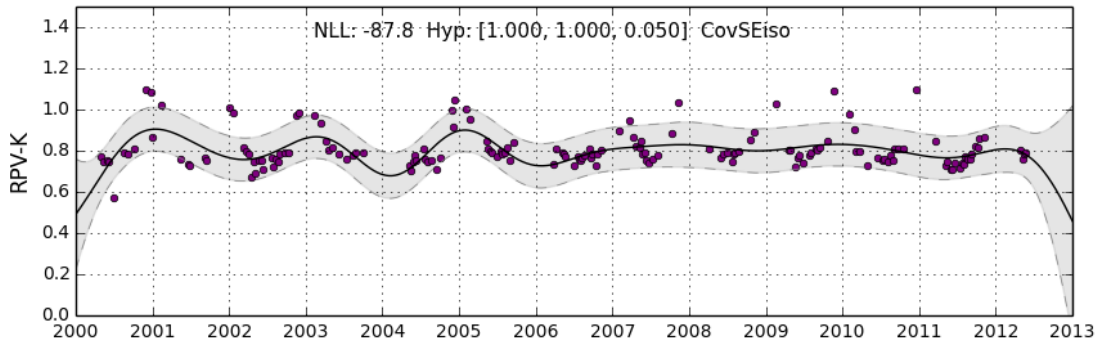

Figure 4.2: Smoothed Time-Series - CovSEiso with Defaults

Figure 4.2 shows a smoothed plot for RPV-K. The curve is the mean prediction of the model for the values. The curve is actually a dense collection of samples from the model for 1000 predicted values

---

[1]The cost function may be a bit conservative as some excluded values could be reasonable, when comparing them to neighboring values. It would be interesting to further research the RPV and TIP cost functions in terms of their outlier predictions.

between the years 2000 and 2013. See Chapter 2 regarding time scale mapping from the model time-scale values to calendar dates.

In terms of the Gaussian Process definitions in Chapter 2, the plot represents equation 2.4, the predicted mean, equation 2.6 and the variances are from equation 2.7. The gray area and the upper and lower dotted gray curves represent an uncertainty band of two standard deviations for the predictions.

The caption at the top of the plot shows the negative log likelihood (NLL), which is the measure for goodness-of-fit. The lower the value, the better the fit (the lower the overall density in the gray band).

The covariance kernel used is also shown in the caption, with the hyper-parameters for the model. In this case, the first two are "effective length" (LS) and the "signal variance" (SV) factors. The last hyper-parameter is always the overall model uncertainty or noise factor (Lik).

As can be seen, this is not a satisfactory model, as the prediction curve is too far from many of the points and predictions between observations would not be accurate.

4.1.3 CovSEiso Kernel - Optimised Hyper-Parameters

The next step is to then try optimising the model to see if a better fit can be obtained. This is done using the GP learning function described in Chapter 2.
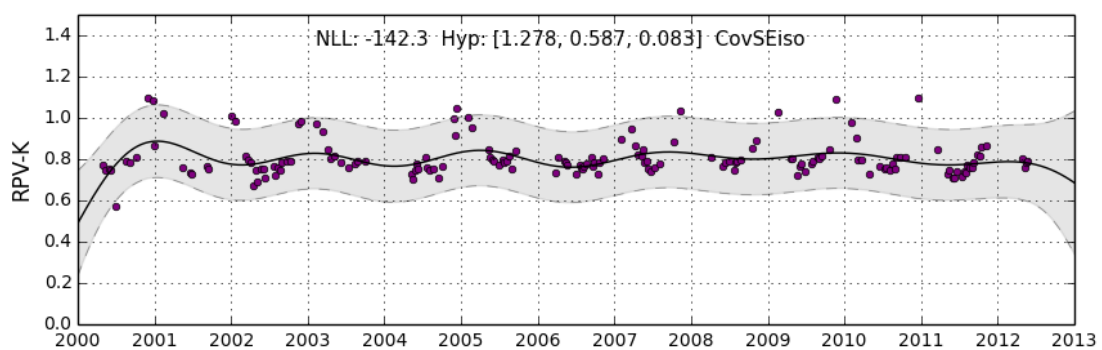

Figure 4.3: Smoothed Time-Series - CovSEiso Optimised

Figure 4.3 shows the optimised model, which although the NLL has improved, is less satisfactory. It seems the data are pushing the CovSEiso kernel towards a linear fit.

4.1.4 CovMaterniso Kernel - Optimised Hyper-Parameters

We then consider an alternative kernel, CovMaterniso which is more suitable for "rougher" data series.
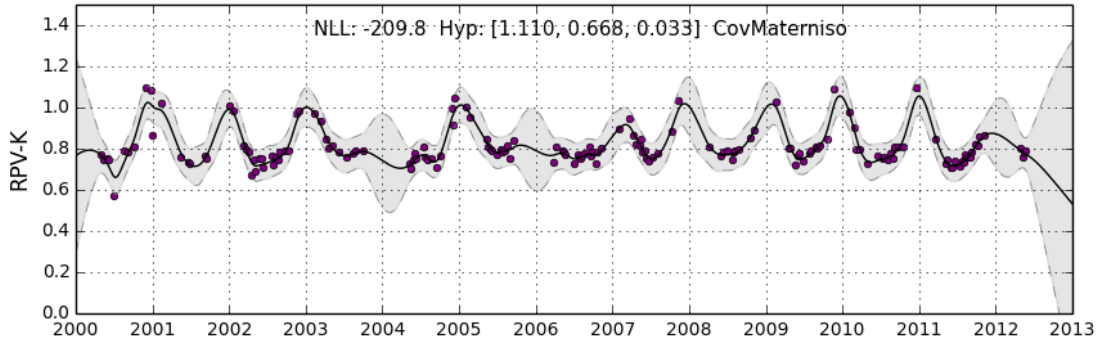
Figure 4.4: Smoothed Time-Series - CovMaterniso Optimised

Figure 4.4 shows the optimised CovMaterniso model, which now provides a reasonable fit for the data and could be used for predicting values across most of the time-series with a reasonable variance band. The overall improvement in the fit can also be seen in the significantly improved NLL value. However, the band widens in places where there are large gaps in the data (typically during the wet season). This clearly shows how gaps in the observations still likely result in inaccurate predictions with the increased uncertainty in those regions.

The optimised hyper-parameters also provide useful information. The effective length factor is 1.111 which shows the "decay" of the covariance effects between observations over their distance apart in time. This also suggests a possible overall periodic structure with a period of around 1.1 years. Similarly, the second hyper-parameter shows a 0.668 decay in the "vertical" covariance influences and suggests a possible amplitude fluctuation of around 0.7 in the RPV-K function space. The overall model uncertainty hyper-parameter of 0.033 gives the indication of a good fit with a low uncertainty over the full joint distribution of the model, but there are some problems with the gaps.

Also noticeable is that the variance band beyond the extents of the observation points rapidly widens, resulting in much higher uncertainty and clearly inaccurate predictions in those regions. Beyond the evidence of the observations, the prediction curve reverts to a sampling from the prior distribution influenced only by the kernel covariances.

4.2 Periodic Structure and Forecasting

Depending on the aims of the analysis, the previous smoothing model may be adequate, or we could attempt to discern the periodic structure in the data series and use that to possibly make better predictions in the gap regions and outside the boundaries of the observations.

4.2.1 Composite Kernel - CovMaterniso and CovPeriodic Summed

The CovPeriodic kernel can produce a periodic model, but with regular periods. We know (a priori) that there is periodicity in our data, but almost certainly with some irregularity in the periods (seasons).

A property of covariance kernels is that they can be combined, through summing kernels together and taking products over kernels. Knowing that the CovMaterniso kernel already has a better fit to the irregularity in the data than the CovSEiso kernel, we will combine the CovMaterniso and the CovPeriodic kernels by summation in order to generate our composite periodic time-series.
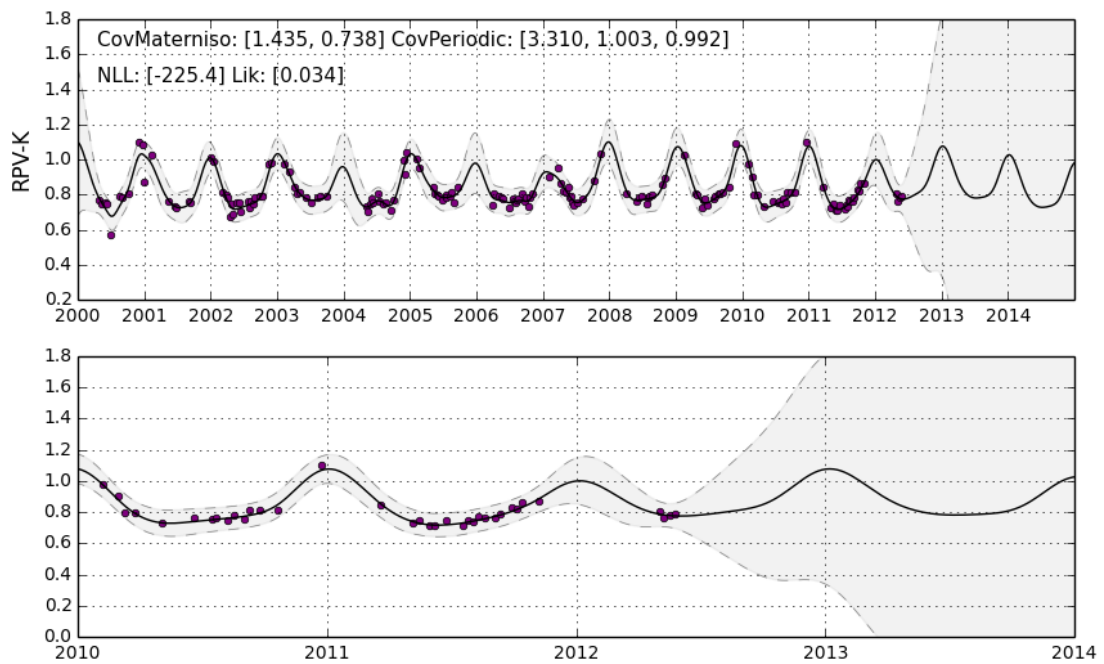
Figure 4.5: Composite Kernel Periodic Time-Series - Initial Optimise

Figure 4.5 shows a strongly periodic model with a much better fit for the gaps and for predicting beyond the boundaries of the observed data. The period has settled to almost exactly one year. The optimisation method used for this model was based on the COBYLA algorithm described in Chapter 2. It is a derivative-free algorithm optimising through internal non-linear inequality constraints. The optimisation module has a "stopping" parameter, which in this case was set at 200 iterations.
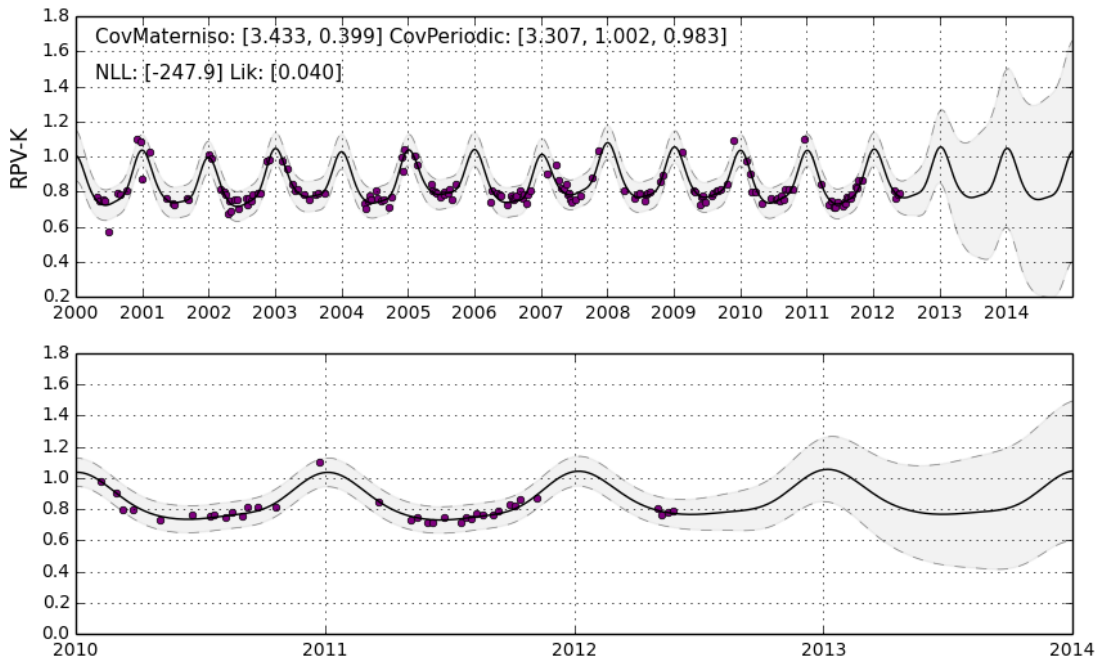
Figure 4.6: Composite Kernel Periodic Time-Series - Optimised

Figure 4.6 shows the model after 1200 iterations with the hyper-parameters reasonably well optimised.

The next question to consider is how the model performs prior to the start of the observations and also the trends over a much longer period, well beyond the previously considered boundaries.
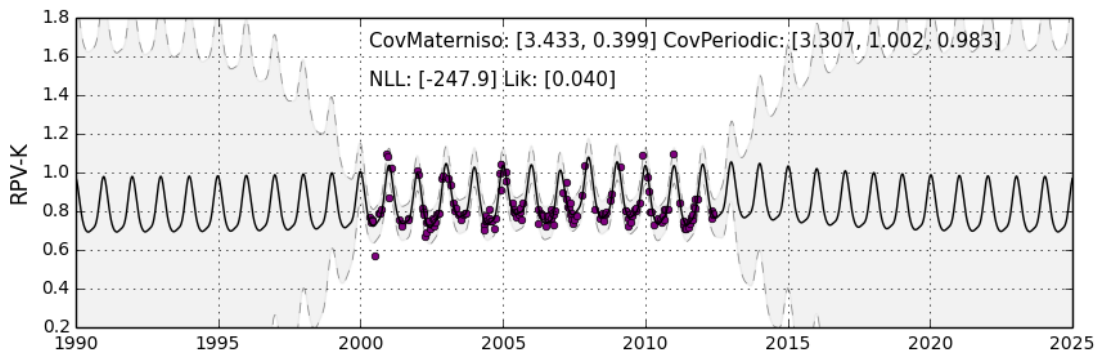

Figure 4.7: Composite Kernel Periodic Time-Series - Long-Term

Figure 4.7 shows the same model posterior (same observations and optimised hyper-parameters), re-sampled on a daily frequence to predict RPV-K values from 1990 through to the end of 2025. The plot shows the classical effect of a Bayesian model: how near the observations, the predictions are strongly "influenced" by the "evidence", while in other regions, the *prior* dominates. In this case the prior comprises both a periodic and an exponential covariance function, and their combined effects in the long term results in a downward drift of the RPV-K predictions to a magnitude of approximately

-0.0015320 between 2024 and 2025 and -0.0014201 between 1989 and 1990. For this particular scenario, this effect would probably be insignificant.

Finally, having obtained a "good" set of hyper-parameters, we can run GP inferencing with the observation data again at any stage to obtain the posterior GP distribution. This can then be used to easily re-sample the model to obtain any desired predictions. The plot in Figure 4.6 was in fact also re-sampled to a daily frequency for the plot.

As an example, Table 4.1 shows a re-sampling to predict RPV-K for specific dates - the Solstice and Equinox dates for the years 2011 to 2013. The model applied is Composite Periodic with 1200 iterations.

Table 4.1: RPV-K Predictions - Solstice and Equinox Dates, Composite Periodic

Date	X-scale	RPV-K	Two Std Dev
20110320	11.2137	0.850864	0.0898527
20110621	11.4685	0.729461	0.0849378
20110923	11.7260	0.781151	0.0849934
20111222	11.9726	1.027530	0.0940931
20120320	12.2158	0.875154	0.0926025
20120620	12.4672	0.765229	0.0969052
20120922	12.7240	0.803227	0.1388730
20121221	12.9699	1.037970	0.1967480
20130320	13.2137	0.885806	0.2570830
20130621	13.4685	0.766379	0.3210520
20130922	13.7233	0.798893	0.3845420
20131221	13.9699	1.029920	0.4411110

Note[2].

4.2.2 Comparing Smoothing and Composite Periodic over other MISR-HR fields

For completeness, in this section we extend the model to perform the time-series over other MISR-HR fields (described in Chapter 2). The table below shows the fields to be used. We include RPV-K again for comparison with the other fields.

Table 4.2: MISR-HR Fields and Descriptions

Field Id	MISR-HR Field	Description
RPV-K	RPV k (red)	RPV model k parameter in the MISR red spectral band
FAPAR	TIP ABS_VIS	Fraction of absorbed photosynthetically active radiation
EFF-LAI	TIP EffLAI	Effective leaf area index
ALB-VIS	TIP BHR_VIS	Bi-hemispherical reflectance (albedo) in the broadband (visible) spectrum.
ALB-NIR	TIP BHR_NIR	Bi-hemispherical reflectance (albedo) in the MISR near infrared spectral band.

---

[2]See Chapter 2 for details on mapping from dates to the real-valued X-scale.

The data extracts are still from the same site, E1 (from Chapter 5).

For purposes of comparing how the CovMaterniso and the Composite Periodic models do when applied to the other MISR-HR fields, Figure 4.8 shows the optimised CovMaterniso model for all the fields, and Figure 4.9 shows the optimised Composite Periodic model (1200 iterations) for all the fields. (Note that the vertical scales are different to the earlier plots due to the compact multiple plot figure format). Table 4.3 shows the hyper-parameters for the CovMateriso models and Table 4.4 shows the hyper-parameters for the Composite Periodic models.
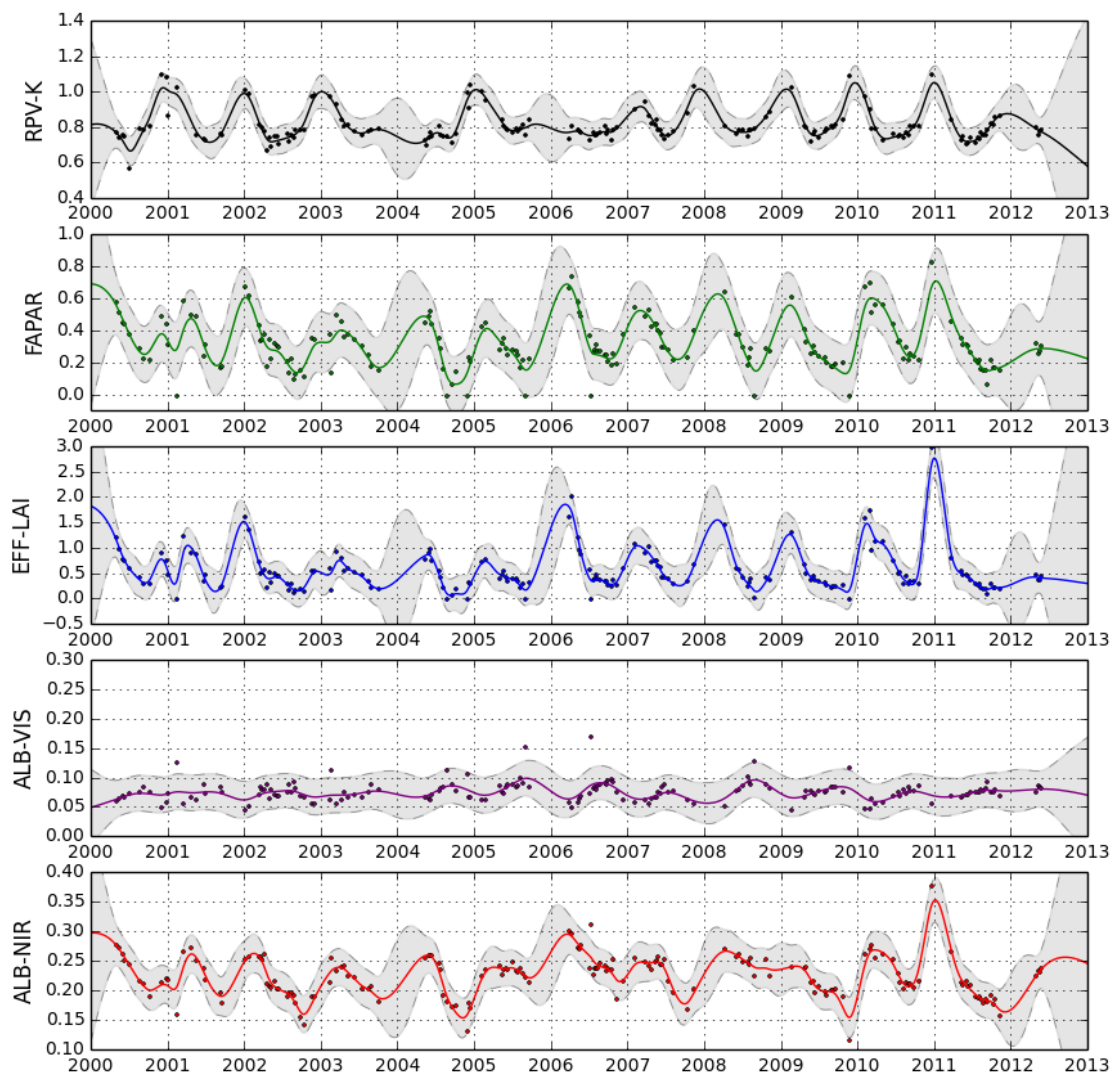

Figure 4.8: Site E1 Full Plate - CovMaterniso.

Table 4.3: Site E1 Full Plate Hyper-parameters - CovMaterniso

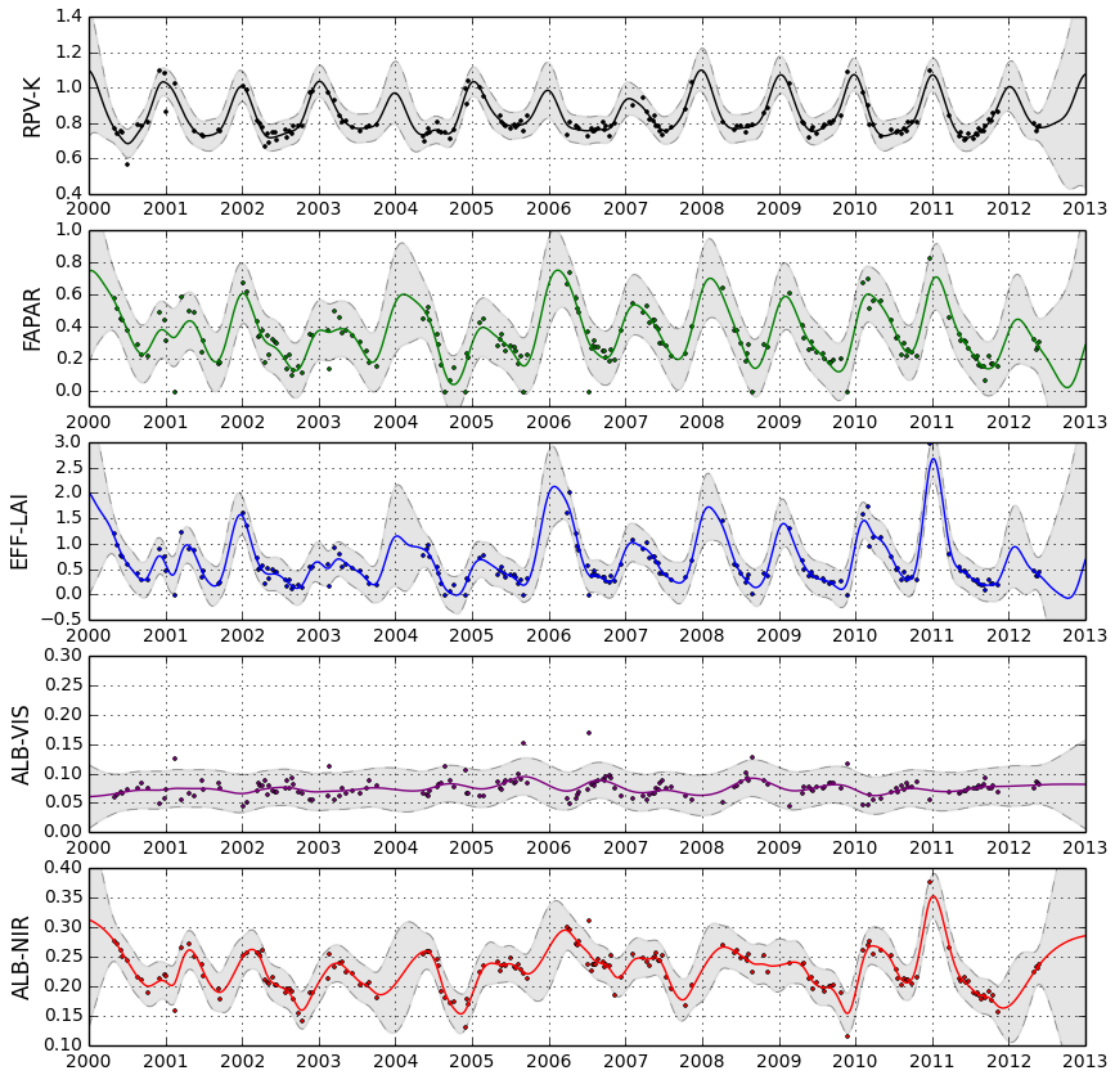| Field Id	NLL	LS_{mat}	SV_{mat}	Lik
RPV-K	-209.186	1.406	0.879	0.034
FAPAR	-105.661	1.125	1.026	0.074
EFF-LAI	41.988	1.133	3.980	0.134
ALB-VIS	-439.335	1.623	0.100	0.014
ALB-NIR	-379.261	1.541	0.370	0.014

Figure 4.9: Site E1 Full Plate - Composite Periodic.

Table 4.4: Site E1 Full Plate Hyper-parameters - Composite Periodic

Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-225.405	1.435	0.738	3.310	1.003	0.992	0.034
FAPAR	-115.887	1.165	0.825	3.487	1.011	0.782	0.076
EFF-LAI	36.468	1.196	2.976	1.712	1.011	0.781	0.147
ALB-VIS	-431.534	1.642	0.133	4.052	4.601	0.608	0.013
ALB-NIR	-376.467	1.523	0.393	3.825	4.740	0.962	0.013

4.3 Conclusions

The Gaussian Process modelling steps carried out in this chapter have each increasingly improved the fit of the models to the data. The final models have delivered a pleasing result for this particular dataset. Reasonable predictions could be made at any point in the time-series regardless of gaps in the data. The time-series could be re-sampled at any desired frequency and predictions made beyond the bounds of the observed data series. Noise or uncertainty in the observations was catered for inherently by assuming a Gaussian distribution over the underlying or latent functions generating the observations. Uncertainty was further catered for by the overall likelihood of the joint distribution for the full model.

The GP modelling allowed different time-series methods to be tried through the different covariance functions, but all within a single theoretical and algorithmic framework. Consistency was achieved in the procedures for specifying hyper-parameters and in carrying out model learning or optimisation.

The results from this chapter have shown that Gaussian Process methods applied to this sample case from MISR-HR data for the Kruger National Park provide reasonable answers in the affirmative to all of the research questions in Chapter 3.

In the next chapter, we further explore Gaussian Process time-series methods over a broader set of sample cases.

Chapter 5

MISR-HR Data Case Study

This chapter covers the development of a case study from MISR-HR data provided by the South African National Space Agency (SANSA) for the Kruger National Park area in north eastern South Africa. The aim is to follow all the processes involved in generating a temporal picture for a twelve year period, of different sites with known differing vegetation features in and bordering the Park. The objective will be to generate full Gaussian Process Regression models for the main MISR-HR fields for each of the sites. As this dissertation is primarily about the generating models for time-series and vegetation phenology rather than a study of the vegetation itself, the objective will be to rather showcase the variability in the MISR-HR data and to highlight some of the obvious temporal patterns from the models. The research objective of this chapter is to further support the thesis of Chapter 3 on broader sets of data than the single set modelled in Chapter 4.

Seven sites have been chosen for this data case study and their particular features described below. The processes from the background laid out in Chapter 2 will be applied, including selecting and geolocating the sites, determining the relevant satellite orbit paths covering the area and from those the data sets and fields to be processed and GP models developed.

5.1   Kruger National Park Sites Chosen For GP Models

The Kruger National Park comprises a wide variety of plant species over habitats varying in geological formations, soil types and climatic conditions. These have been grouped into twelve vegetation landscape regions covering the overall area of the park [Venter and Gertenbach 1986], [Joubert 2007]. Our data study will be constrained to the north western portion of the park, to an area in which four of the vegetation landscapes are found. One site from each of these will be selected, to show the temporal patterns for each of the MISR-HR variables for each of these sites, with plots facilitating comparisons between the sites.

There are also many fire events which occur in the park, both managed and natural. Some fires are known to have occurred in this region during 2010[1], within two of the vegetation landscapes, so two further sites are chosen where these fires have occurred in order to see whether these events are evident in the models. Finally, human development has occurred outside the boundary of the park bordering the area near the

---

[1]Fire activity was observed in the MISR-HR images while the data were being explored for preparation of the case study.

above-chosen sites. One further site has been chosen in this area, also to see if distinguishing patterns are evident in the models.

5.2   Modelling Methodology

The methodology for the data case study is as follows. The Global Grid (longitude and latitude) coordinates for the study area are determined [Joubert 2007]. The MISR orbit paths and blocks covering the area are identified in order to determine which MISR-HR datasets apply.

Visual images of the blocks are generated to confirm the geolocation and determine overlapping of the blocks for the area and thus which datasets contain the required data.

A more detailed visual image is then generated in order to identify target study sites in each vegetation landscape, the areas affected by fire and the geolocation of a site bordering the park in the area.

The data for the sites are then extracted and GP models are generated using the Composite Periodic covariance kernel described in Chapters 2 and 4. Plot "plates" showing the full historical models for each site are produced. Further plots are also produced allowing comparisons between the sites.

5.3   Geolocation, Block and Dataset Selection

The MISR-HR data for the Kruger National Park provided by SANSA, comprised datasets from the orbit paths 169 and 170 and the blocks 109 and 110 in both paths. The northern part of the park is covered by block 109 and the southern part by block 110. For this study the datasets used are from block 109 and will be referred to as 169/109 and 170/109 to distinguish the two orbit paths for the block.

Figure 5.1 shows a schematic of these and their longitude and latitude positioning. The figure shows the extent to which the blocks overlap.
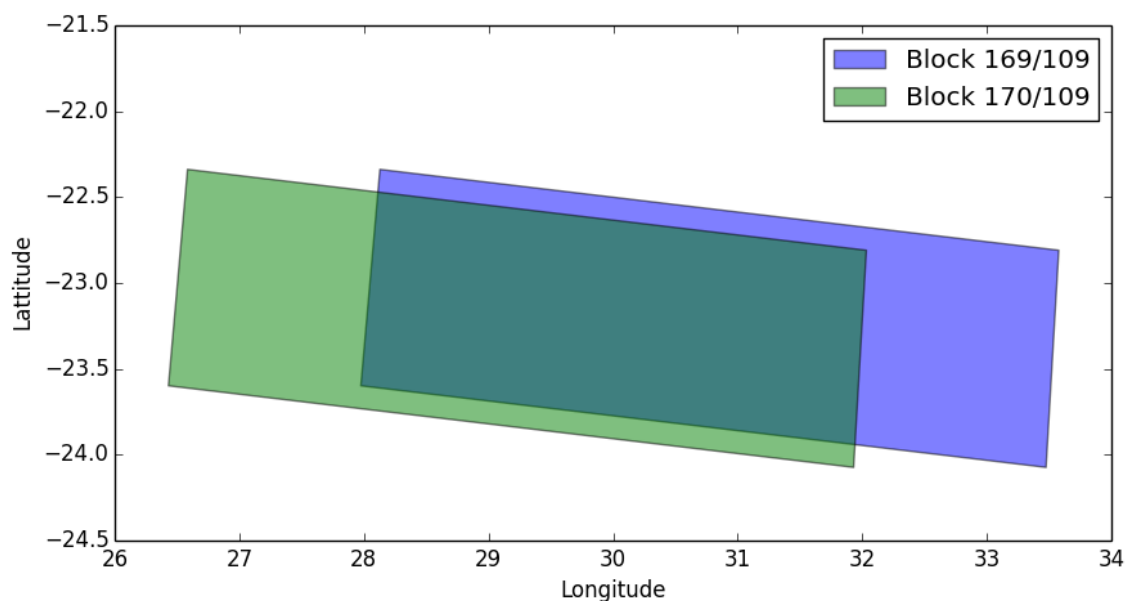

Figure 5.1: Overlapping Orbit Blocks.

The next step was to generate images of the blocks 169/109 and 170/109 from the MISR-HR datasets from orbits of the 18th August 2010 and 9th August 2010 respectively. The Eos.jl utility described in Chapter 2 was used to extract the surface Bidirectional Reflectance Factor (BOA-BRF) data for the nadir camera for the blue, green and red spectral bands. A "natural" pseudo-colour map was applied and the BRF data composited into standard RGB images using the Plot.jl component. These can be seen in Figures 5.2 and 5.3.

The blocks are each 512×2048 pixels with the axes representing the relative SOM block coordinates. The verticals are referred to as the "lines" and the horizontals the "samples". The block thus comprises 512 lines and 2048 samples. Individual pixels represent a 275 m square area on the surface and their positions are referred to as line-sample points. For example, the top left corner is at line-sample point (0,0) and the bottom right at point (511,2047).
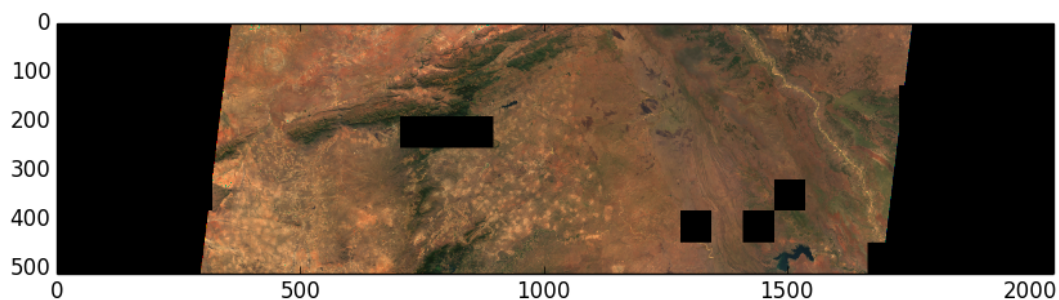

Figure 5.2: Composited MISR-HR Image for Block 169/109, 2010-08-18.

The black regions on the left and right sides of the images are unusable "edge" pixels. The black blocks within the images are 17.6 km by 17.6 km areas where in standard MISR, the characterisation of the atmosphere did not converge and provide a "good enough" description of the aerosol properties to derive a reliable surface product. Causes could be: presence of clouds; bright surfaces; highly variable topography; or other factors affecting the atmospheric processing module in MISR.

In Figure 5.2, the northern part of the Kruger National Park can be seen falling roughly between samples 1100 and 1350. To the right of 1350 the Lebombo mountain range separates South Africa from Mozambique with the Limpopo river visible on the Mozambique side.
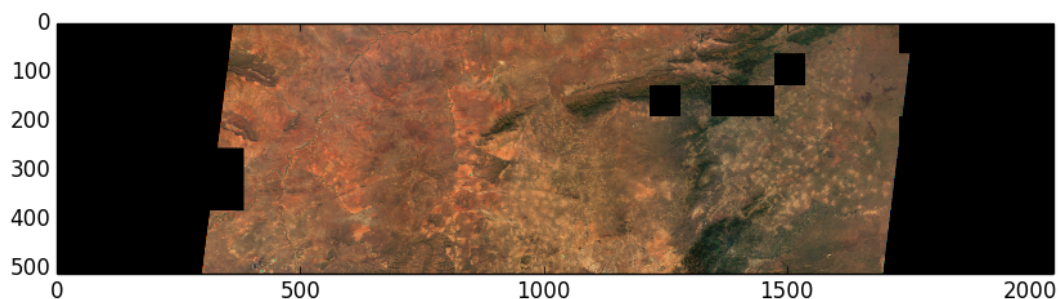

Figure 5.3: Composited MISR-HR Image for Block 170/109, 2010-08-09.

In Figure 5.3, the western boundary of the park is roughly at sample 1650, with the eastern part of the park obscured as edge pixels. However, the study area falls within the visible portion[2].

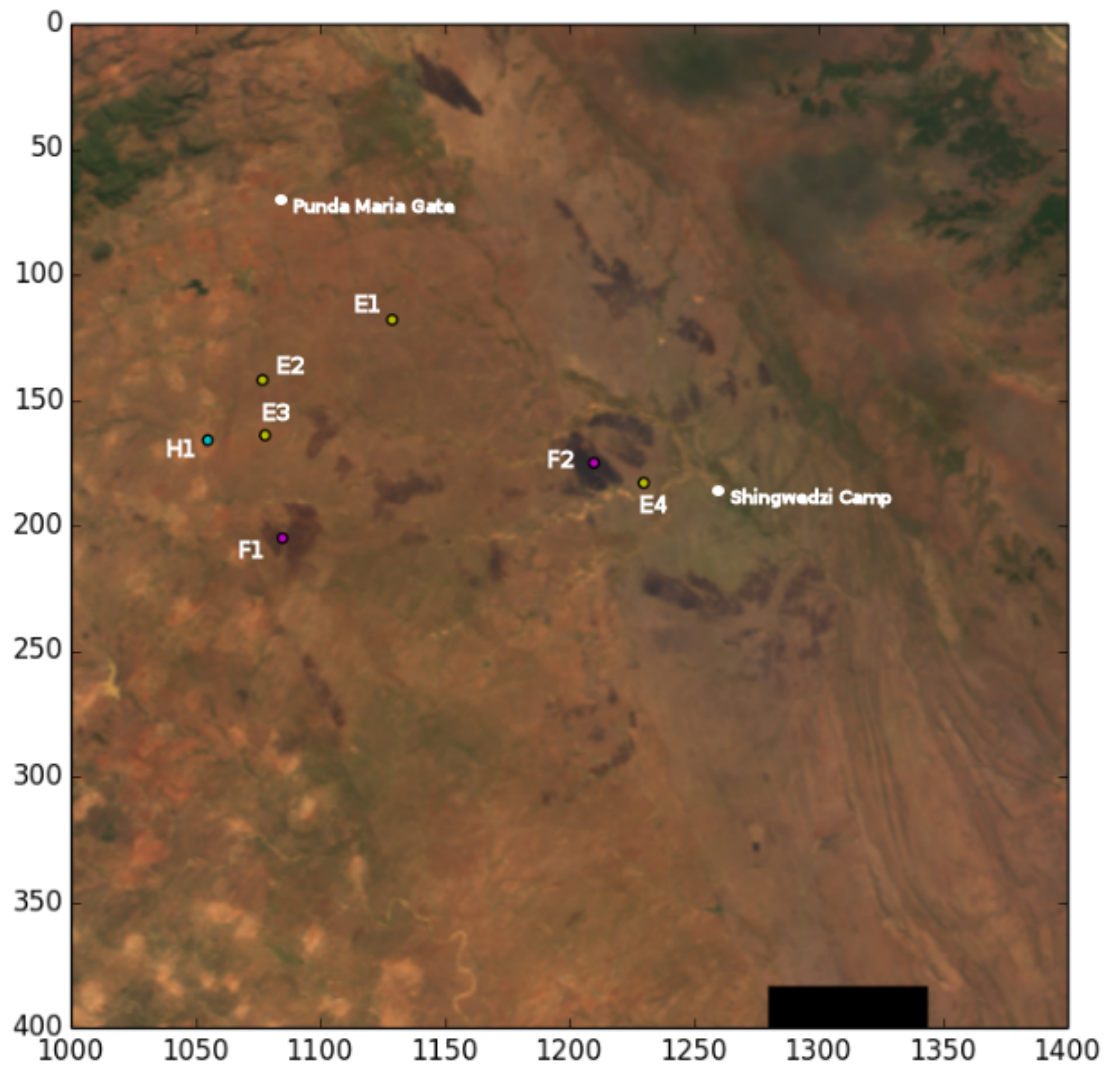## 5.4 Target Site Identification

Figure 5.4: Image Showing Study Area and Target Sites.

[2]Block 168/109 would also have been suitable, but this was not available at the time of obtaining the data.

The image in Figure 5.4 shows a magnification of block 169/109 to view the area for the study in more detail. The image is from the MISR-HR dataset from the orbit of August 18th, 2010.

Two known locations, Punda Maria Gate and Shingwedzi Camp, were geolocated from their longitude and latitude coordinates. This was done using the Eos.jl utility to convert from global grid coordinates to SOM line/sample points. They are marked at the appropriate points on the image.

Target sites for the four landscape regions were selected by referring to landscape maps for the Kruger National Park from [Joubert 2007]. Eos.jl was used in this case to convert from SOM to global coordinates as the sites were chosen from line-sample points on the image.

The "patchy" areas to the west of the sites show the region bordering the Park in which human development is evident.

The following table shows the site identifiers, landscape names and their global and line/sample coordinates.

Table 5.1: Study Target Sites

ID	Landscape	Longitude	Latitude	Line	Sample
E1	Mopane Savannah	31.10083	-22.91293	118	1129
E2	Sandveld Communities	30.95611	-22.96046	142	1077
E3	Mopani/Bushwillow Woodlands	30.95336	-23.01509	164	1078
E4	Mopani Shrub Veld	31.25602	-23.88052	183	1193
F1	Fire Site 1	30.96195	-23.11807	205	1085
F2	Fire Site 2	30.95611	-22.96046	175	1210
H1	Human Development	30.89142	-23.01477	166	1055

5.4.1 Data Extraction and Processing

The data extraction and transformation utility Eos.jl was used for extracting the history data for the target site points from both blocks 169/109 and 170/109, as these orbit paths happen at different times for the same location and thus increase our density of observations. The two sets of observations are combined for processing.

5.5 Composite Periodic Models For all Sites

Based on the modelling process done for site E1 in Chapter 4, we will use the Composite Periodic model, CovSum (CovMatern, CovPeriodic), to generate the models and plate plots for all the sites. The fields are as described in Table 4.2. Although the plate for site E1 was also shown in Chapter 4, it is repeated here for a complete view of all the sites together.
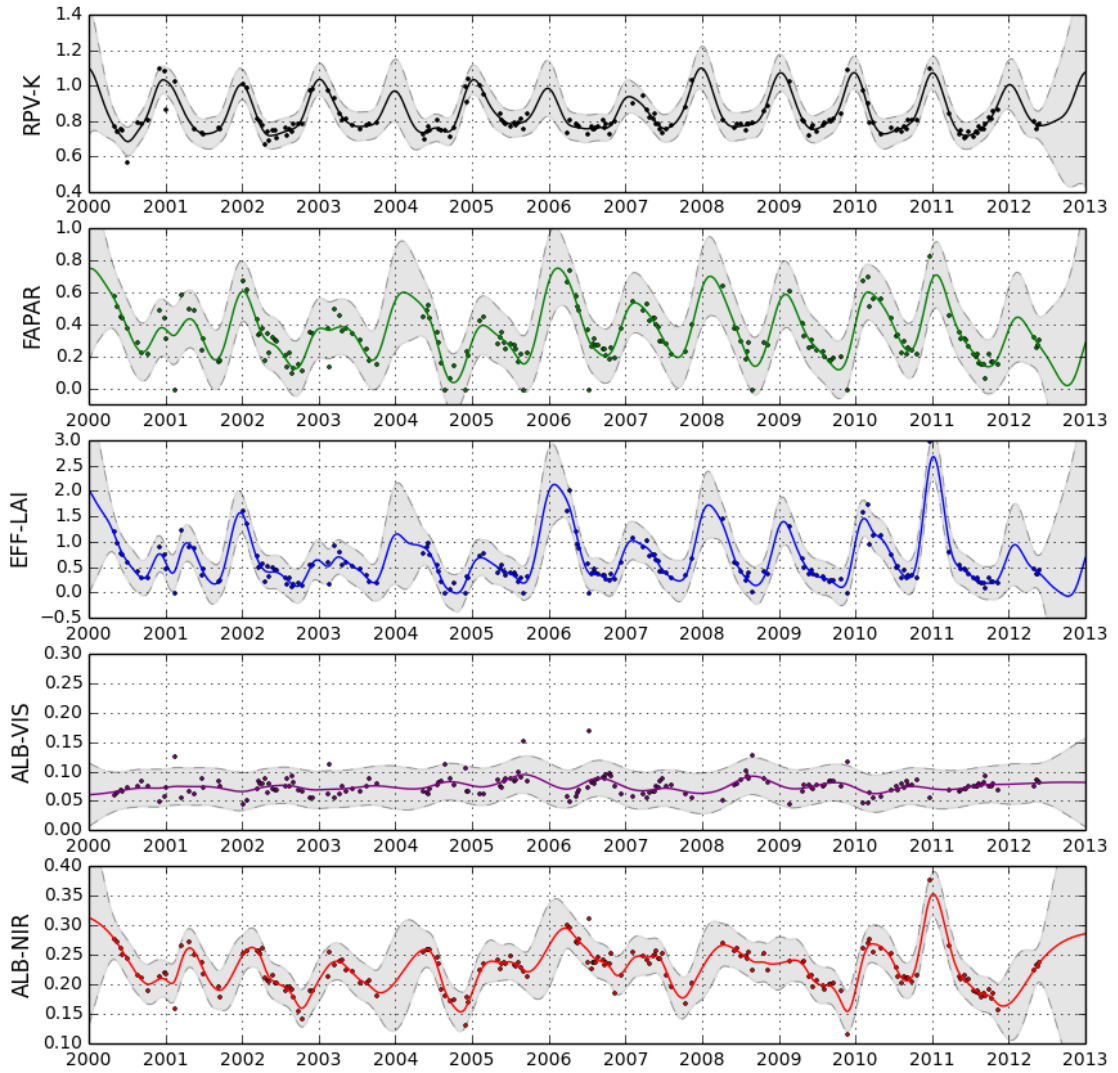
Figure 5.5: Site E1 Full Plate - Mopane Savannah.

Table 5.2: Site E1 Full Plate Hyper-parameters - Mopane Savannah

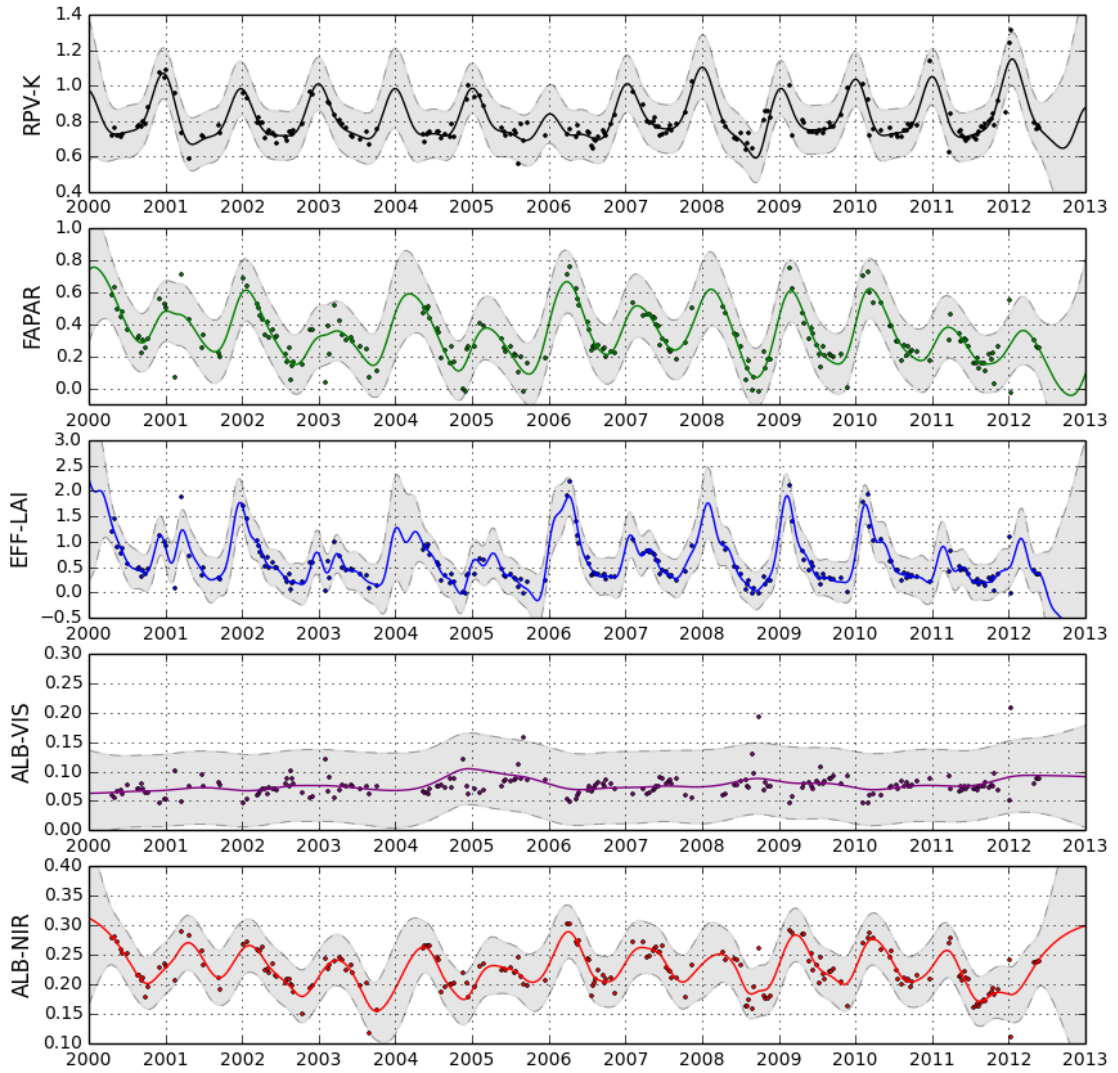| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-225.405	1.435	0.738	3.310	1.003	0.992	0.034
FAPAR	-115.887	1.165	0.825	3.487	1.011	0.782	0.076
EFF-LAI	36.468	1.196	2.976	1.712	1.011	0.781	0.147
ALB-VIS	-431.534	1.642	0.133	4.052	4.601	0.608	0.013
ALB-NIR	-376.467	1.523	0.393	3.825	4.740	0.962	0.013

Figure 5.6: Site E2 Full Plate - Sandveld Communities.

Table 5.3: Site E2 Full Plate Hyper-parameters - Sandveld Communities

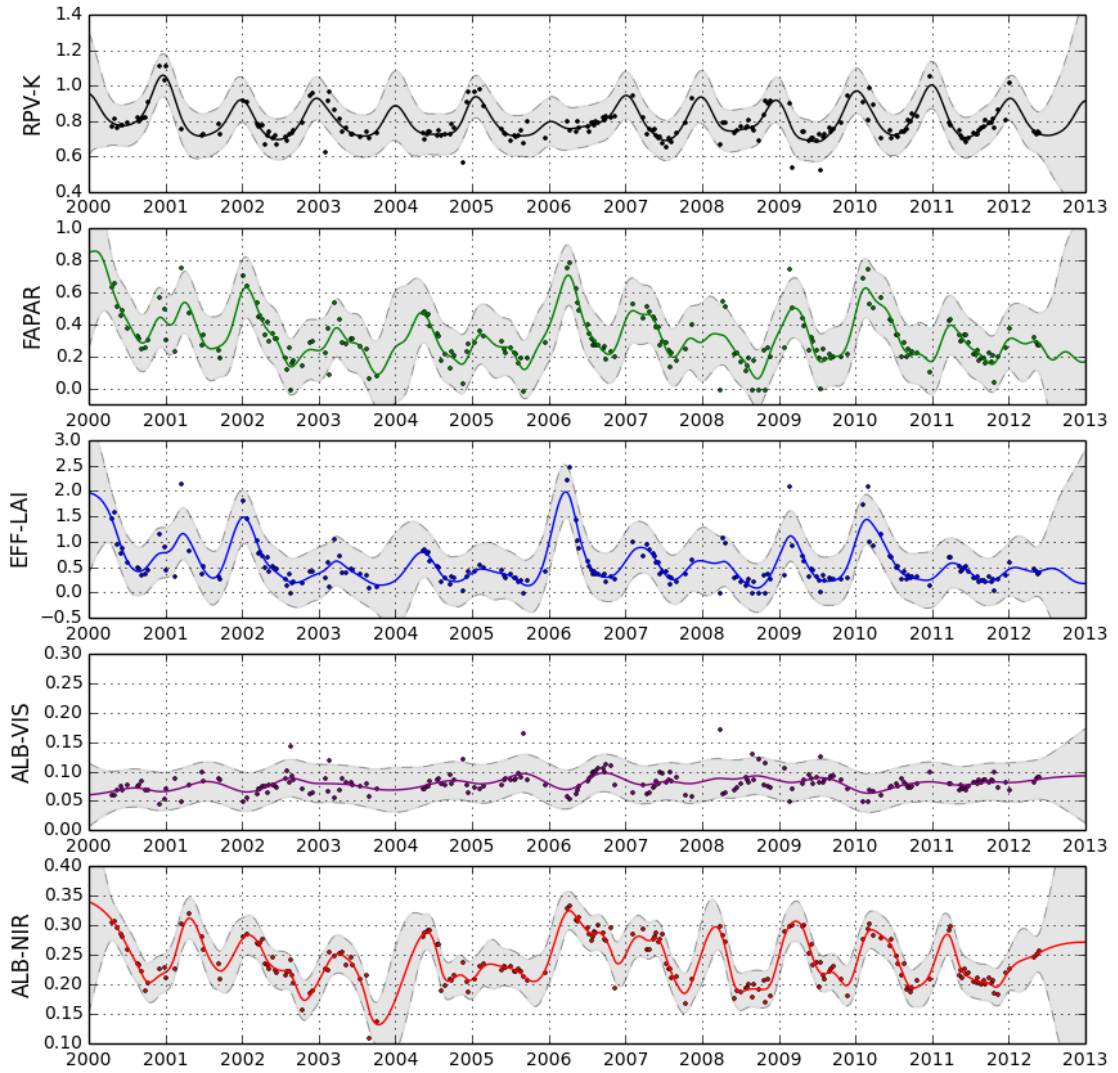| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-179.444	1.500	0.792	3.256	1.004	0.881	0.063
FAPAR	-135.031	1.177	0.752	3.373	1.020	0.944	0.081
EFF-LAI	45.666	1.220	3.215	1.406	1.020	3.899	0.167
ALB-VIS	-372.094	1.421	0.141	4.043	5.153	1.158	0.029
ALB-NIR	-402.966	1.608	0.315	3.323	6.332	0.631	0.020

Figure 5.7: Site E3 Full Plate - Mopani/Bushwillow Woodlands.

Table 5.4: Site E3 Full Plate Hyper-parameters - Mopani/Bushwillow Woodlands

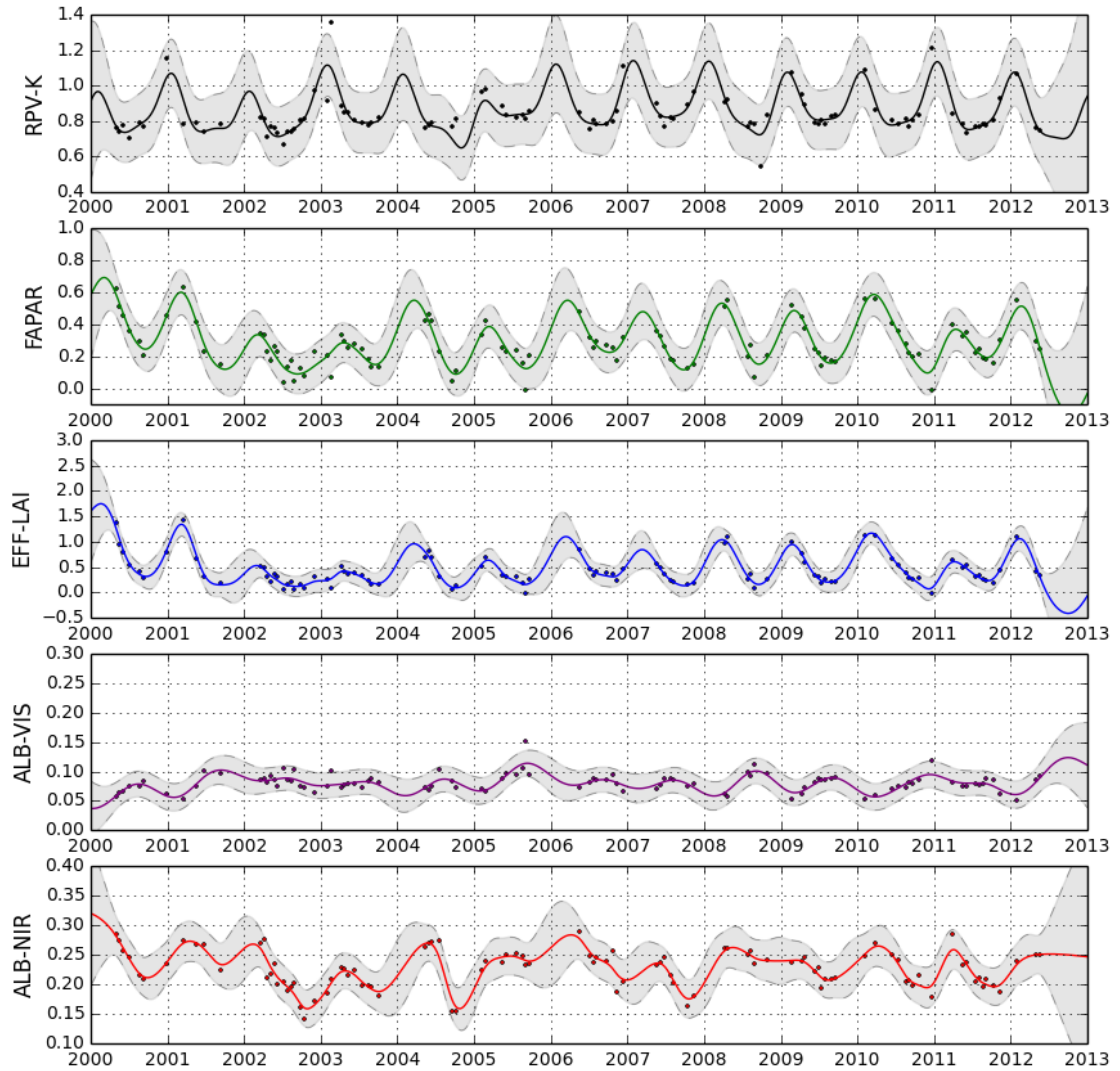| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-213.743	1.436	0.700	3.384	1.003	0.906	0.052
FAPAR	-126.493	1.098	0.852	2.950	0.765	0.542	0.080
EFF-LAI	66.961	1.133	2.225	1.889	1.178	0.539	0.219
ALB-VIS	-487.701	1.573	0.138	4.581	3.578	0.683	0.014
ALB-NIR	-431.458	1.529	0.420	3.590	4.292	0.909	0.014

Figure 5.8: Site E4 Full Plate - Mopani Shrub Veld.

Table 5.5: Site E4 Full Plate Hyper-parameters - Mopani Shrub Veld

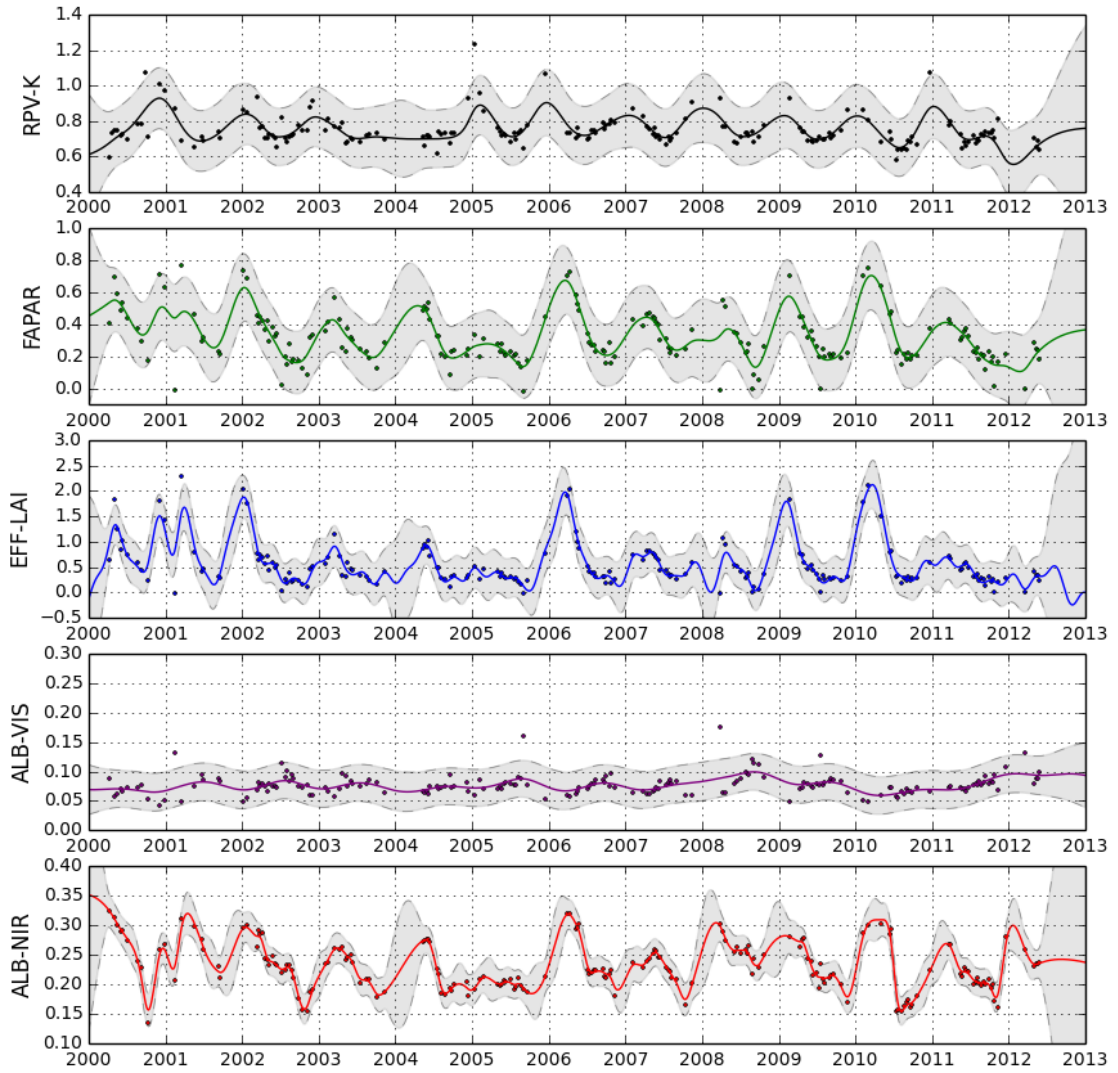| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-57.468	1.441	0.888	3.240	0.996	0.866	0.071
FAPAR	-79.926	1.475	0.807	3.448	1.001	0.869	0.053
EFF-LAI	-11.835	1.209	1.981	3.334	0.999	0.981	0.079
ALB-VIS	-254.306	1.601	0.098	2.145	4.034	1.289	0.009
ALB-NIR	-207.254	1.390	0.210	3.228	7.453	0.814	0.013

Figure 5.9: Site F1 Full Plate - Fire Site 1.

Table 5.6: Site F1 Full Plate Hyper-parameters - Fire Site 1

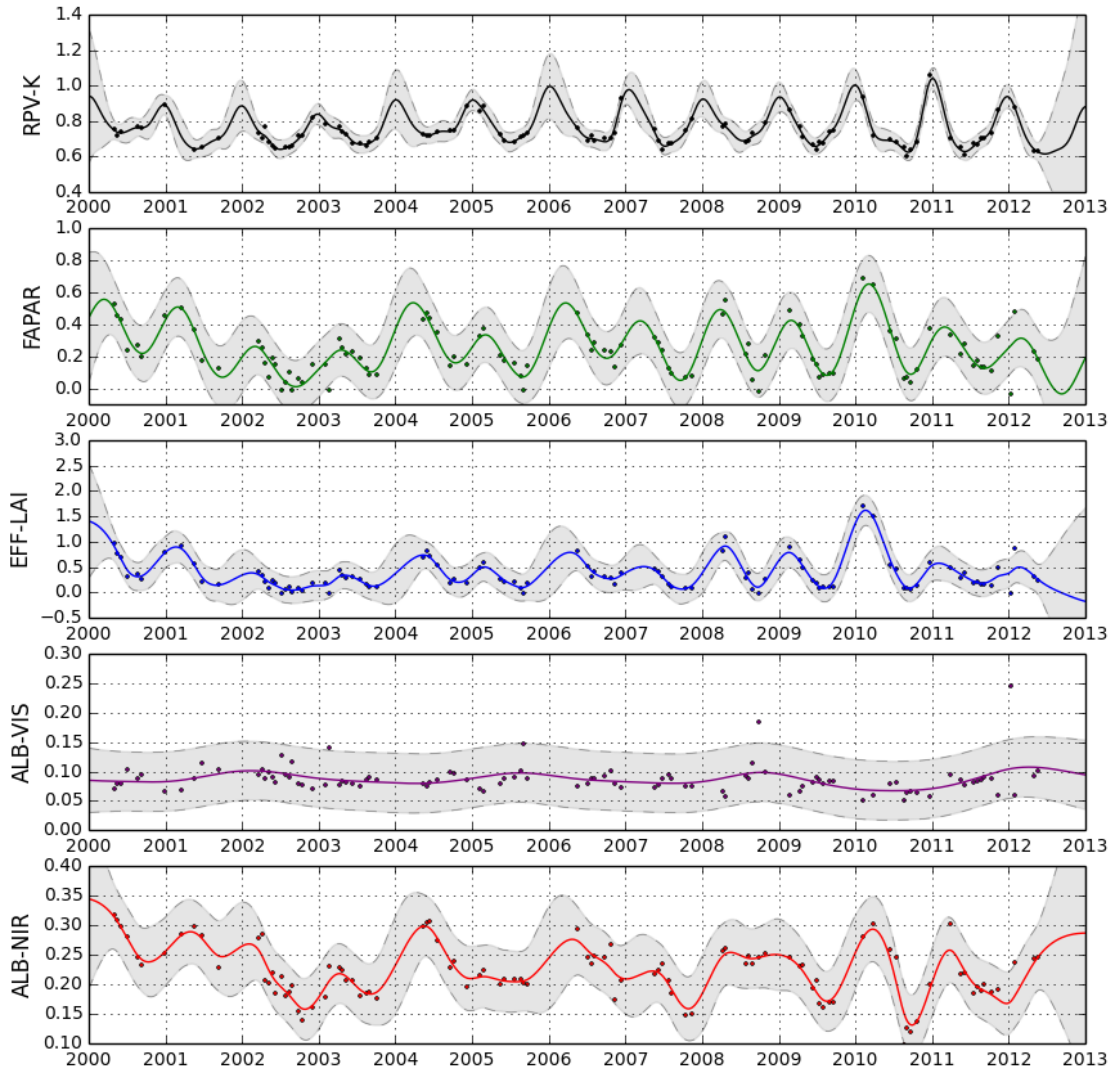| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-151.890	1.190	0.539	3.662	2.678	0.709	0.074
FAPAR	-106.798	1.223	0.927	3.481	3.055	0.795	0.090
EFF-LAI	70.143	1.163	3.627	1.342	1.173	3.670	0.170
ALB-VIS	-472.911	1.492	0.108	3.850	3.157	0.583	0.014
ALB-NIR	-407.078	1.645	0.742	2.956	4.655	0.561	0.010

Figure 5.10: Site F2 Full Plate - Fire Site 2.

Table 5.7: Site F2 Full Plate Hyper-parameters - Fire Site 2

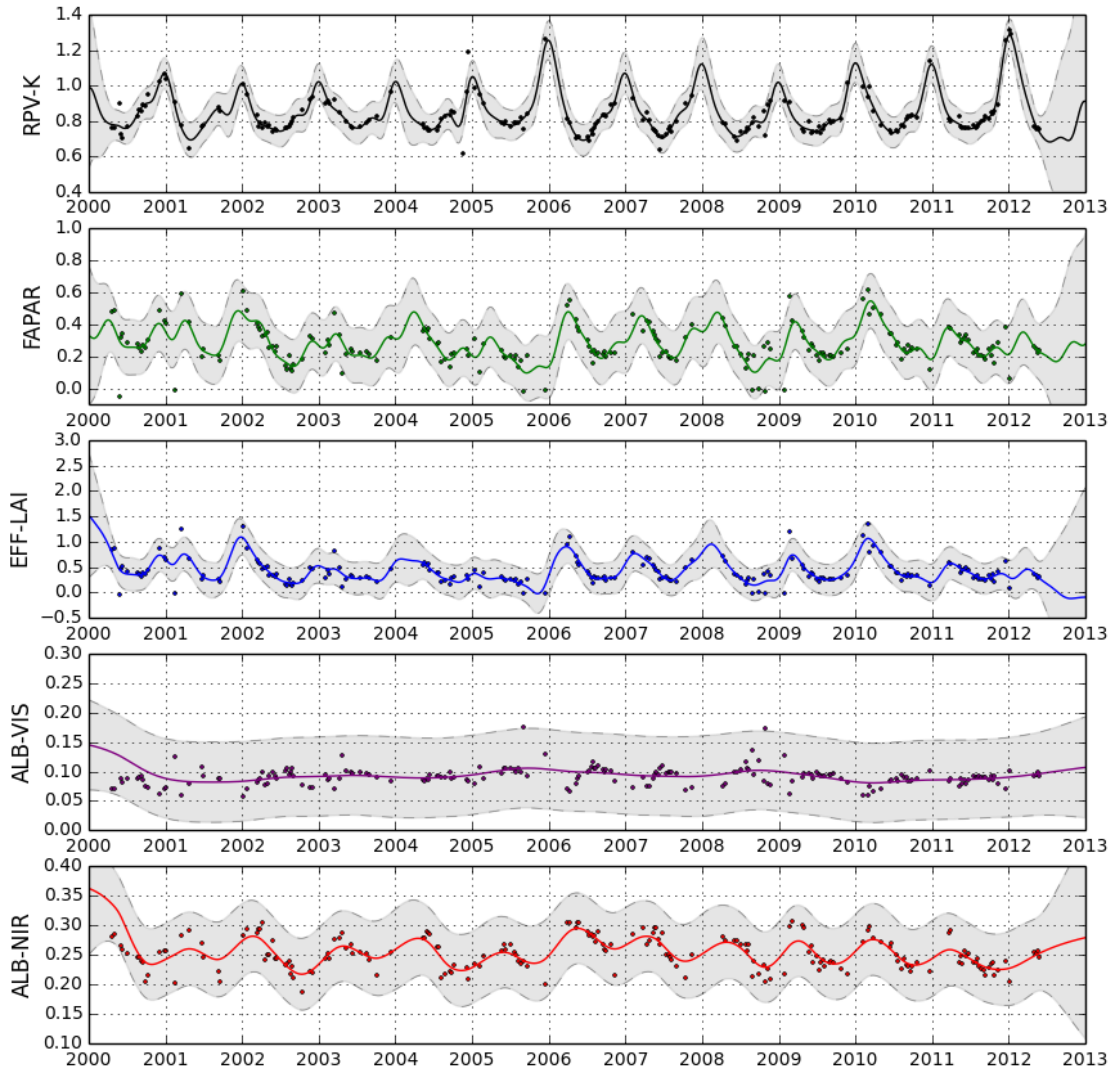| Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-122.718	1.469	0.842	3.309	1.001	0.745	0.018
FAPAR	-66.240	1.483	0.812	3.583	0.997	0.831	0.071
EFF-LAI	5.913	1.218	1.678	3.378	1.262	0.506	0.116
ALB-VIS	-209.775	1.665	0.088	4.012	3.442	0.387	0.023
ALB-NIR	-178.332	1.422	0.247	2.855	6.635	0.449	0.022

Figure 5.11: Site H1 Full Plate - Human Development.

Table 5.8: Site H1 Full Plate Hyper-parameters - Human Development

Field	NLL	LS_{mat}	SV_{mat}	LS_{per}	P_{per}	SV_{per}	Lik
RPV-K	-240.383	1.477	0.959	1.901	0.999	1.035	0.041
FAPAR	-151.980	1.152	0.576	2.151	0.993	0.942	0.076
EFF-LAI	-9.288	1.199	2.461	1.754	1.031	0.902	0.133
ALB-VIS	-346.457	1.637	0.187	4.132	4.716	1.234	0.032
ALB-NIR	-364.351	1.504	0.210	4.078	5.265	0.822	0.028

5.6 Comparing Sites E1, E2, E3, E4 in Different Landscapes

The following are the plots for the sites in different landscapes for each variable, for just the periods from 2008 to 2012. This allows a more direct comparison of the differences between the vegetation landscape sites for each of four MISR-HR variables.
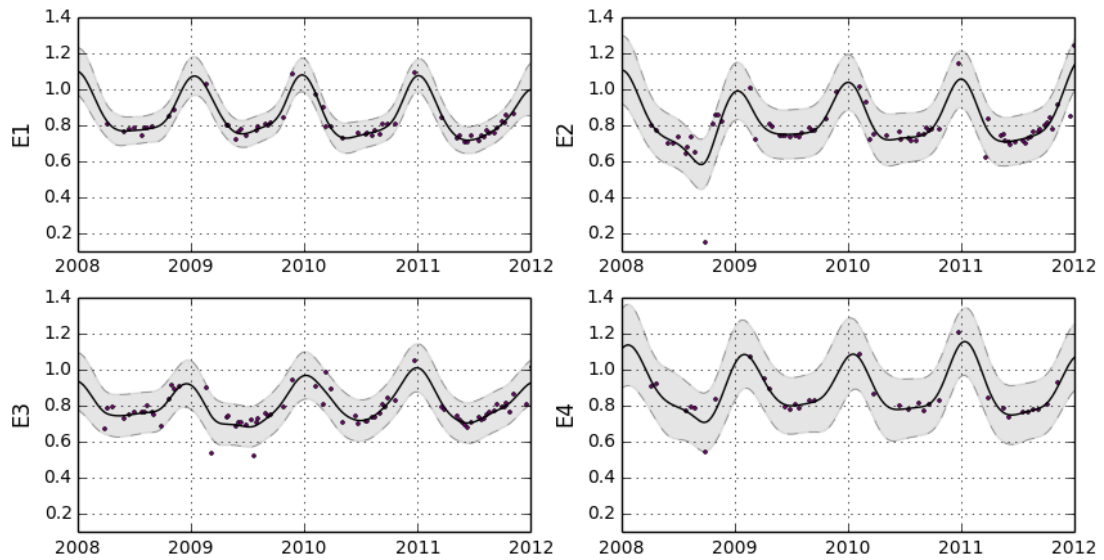

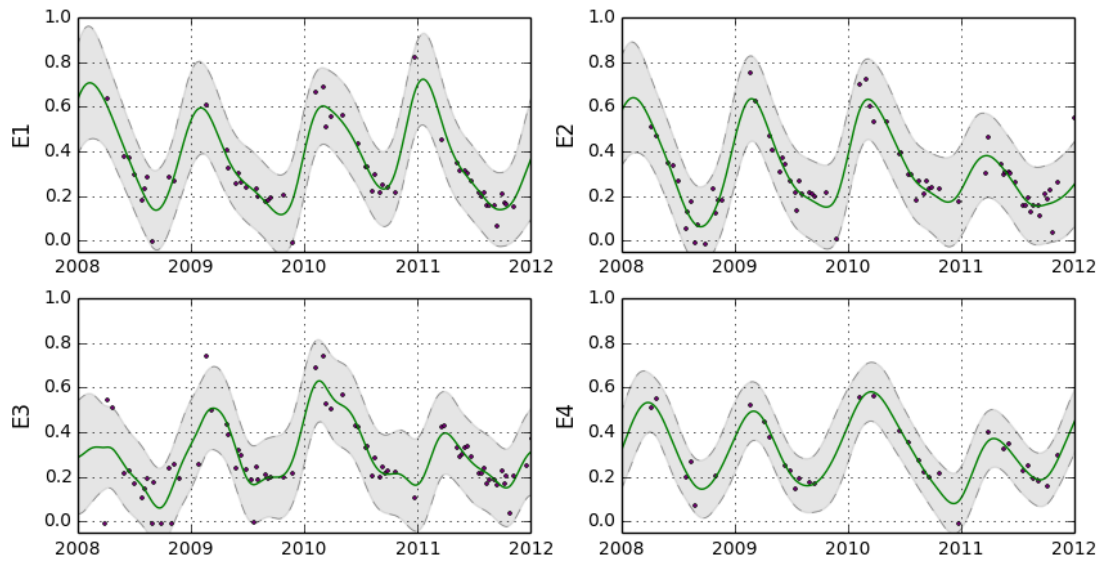
Figure 5.12: Sites E1, E2, E3, E4 - RPV-K.

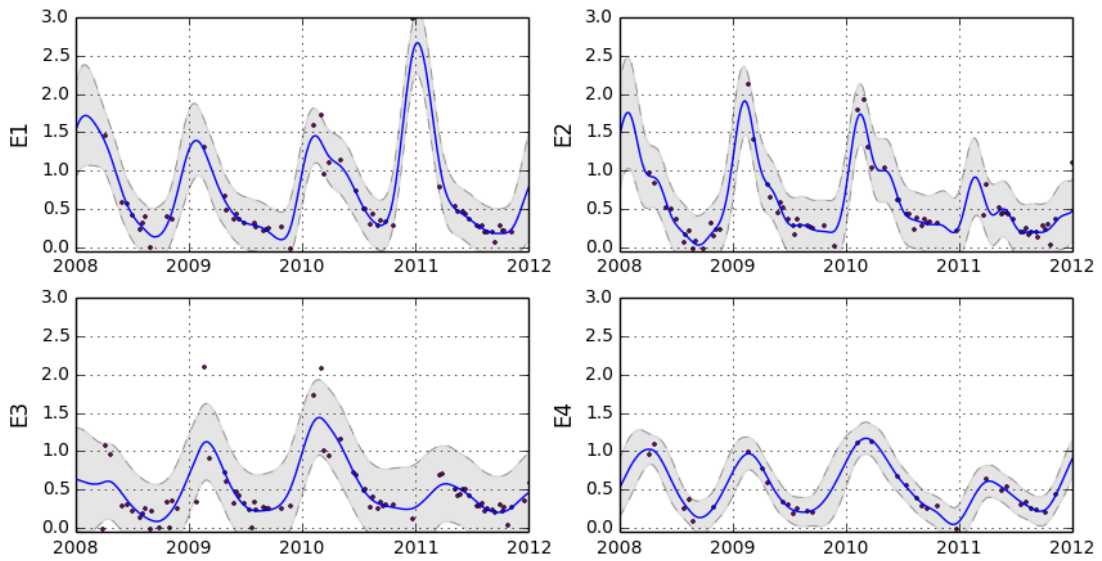Figure 5.13: Sites E1, E2, E3, E4 - FAPAR.
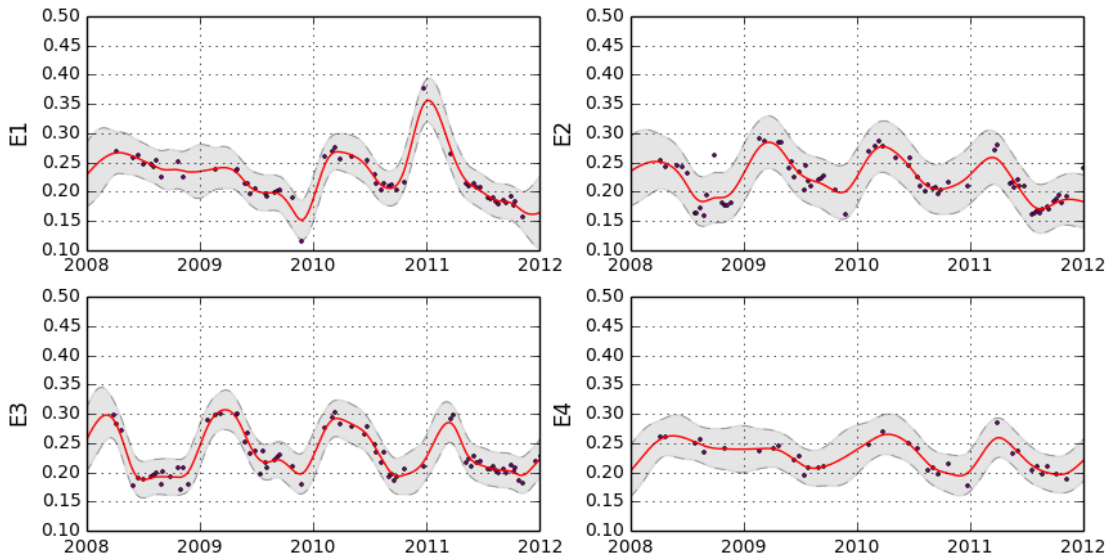
Figure 5.14: Sites E1, E2, E3, E4 - EFFLAI.

Figure 5.15: Sites E1, E2, E3, E4 - ALBEDO-NIR.

5.7 Comparing Sites E3, F1, F2, H1 for Fire and Human Development

The following plots provide the same information for the fire sites F1 and F2, and the bordering site H1. The nearest vegetation landscape site E3 is also included for comparison purposes. Four MISR-HR variables are plotted for just the periods from 2008 to 2012. This allows a more direct comparison of the differences between these alternative feature sites and a vegetation landscape site.
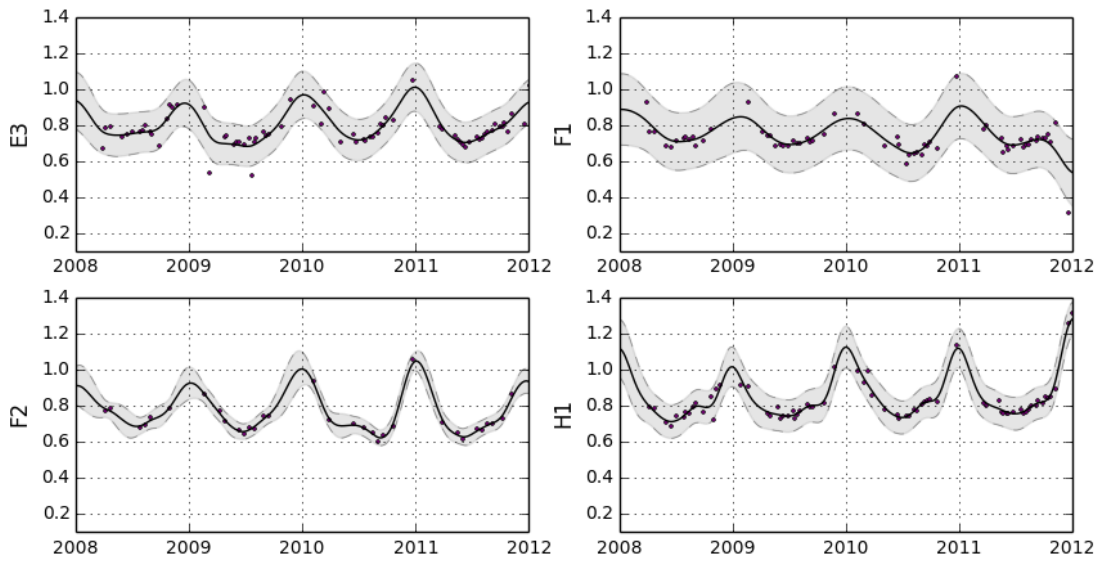
Figure 5.16: Sites E3, F1, F2, H1 - RPV-K.
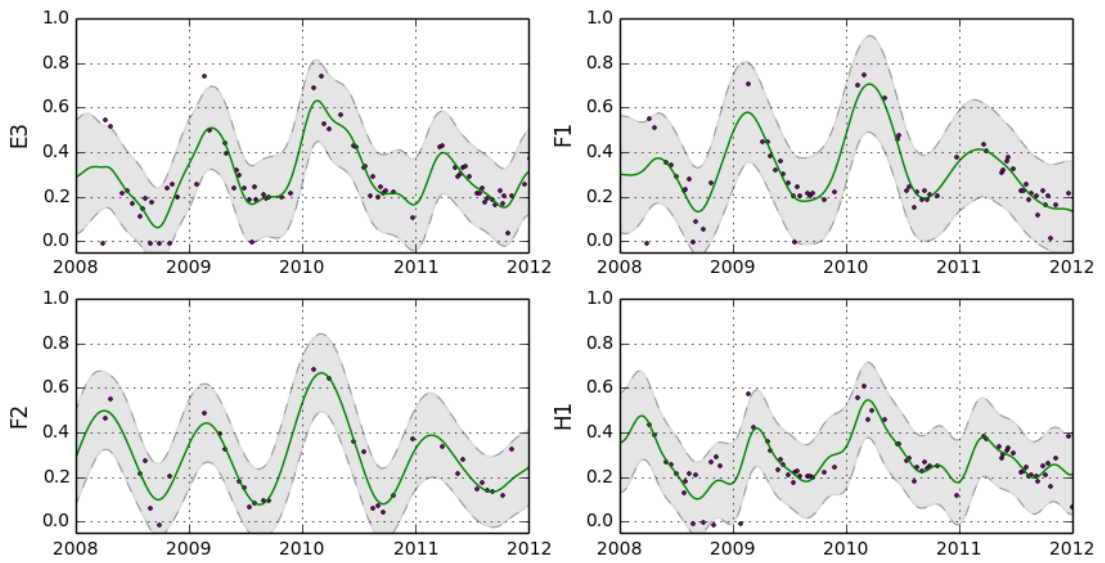
Figure 5.17: Sites E3, F1, F2, H1 - FAPAR.
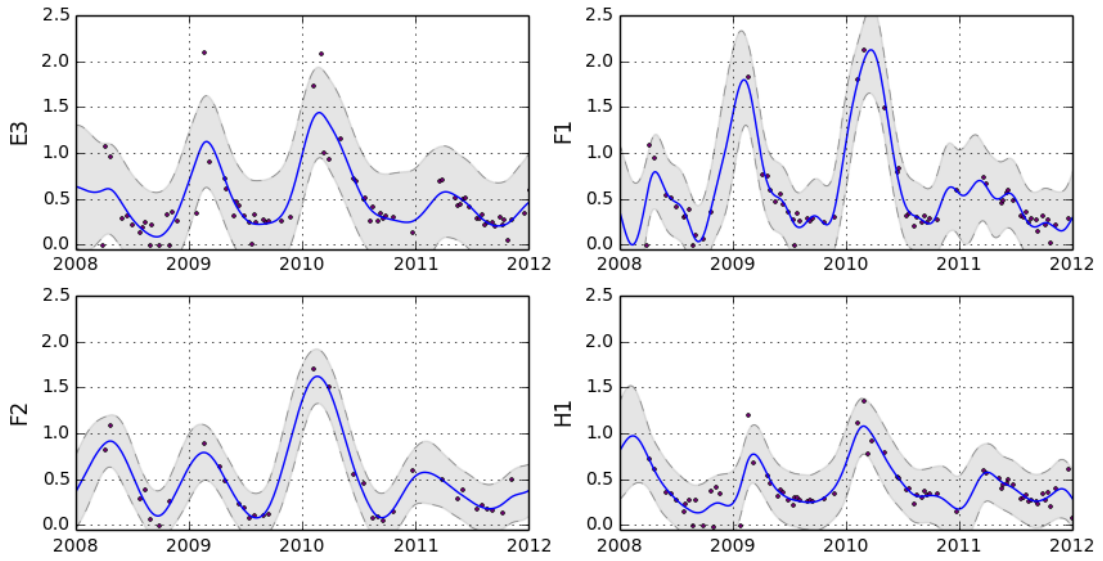
Figure 5.18: Sites E3, F1, F2, H1 - EFFLAI.
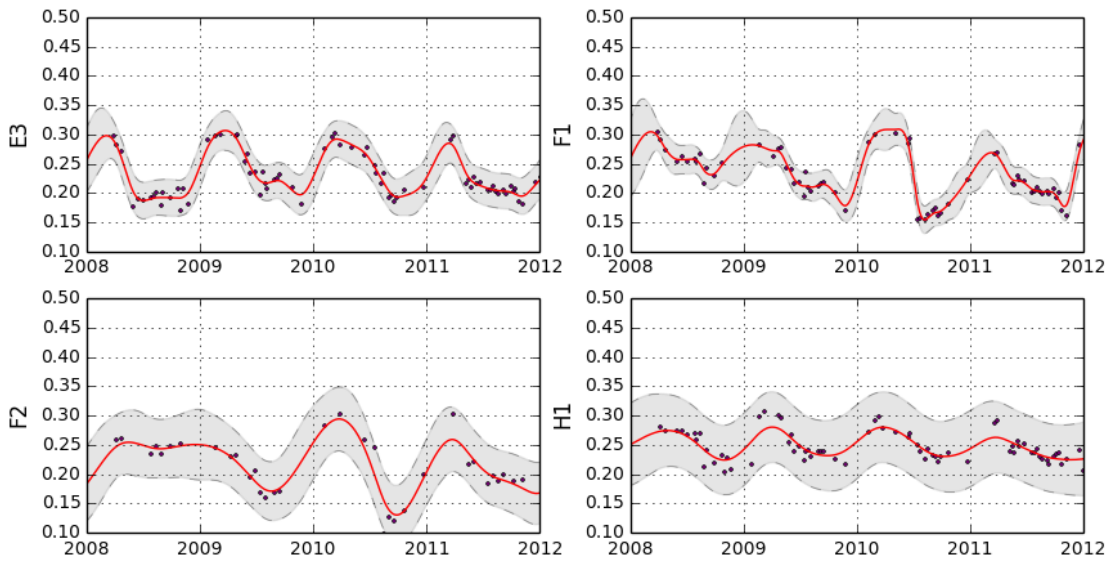
Figure 5.19: Sites E3, F1, F2, H1 - ALBEDO-NIR.

5.8 General Observations From Site Model Plots

These plots (including the more comparative plots from the previous section), suggest the following comments:

The data for Site E4 are somewhat sparser than for the other vegetation sites. That is, there are fewer observations across the time-series than for the other sites. The reason for this is that the site is further

to the east and actually fell within the unusable edge region of block 170/109, thus no data was available from that overlapping orbit[3]. In spite of this, the models still show reasonable results in all the fields and could still be used in further phenology studies. The higher (less negative) NLL values indicate however, the reduced confidence in the overall likelihood fit for the models.

Generally, the NLL values for the EFF-LAI field are poor, indicating that the Composite Periodic model is probably not a good choice for this field. Further modelling could be done to improve this by perhaps using a CovMaterniso kernel with a higher roughness class, for example $V = 1$ instead of the $V = 3$ used above. More complex combinations of kernels also could be attempted. As an example, Figure 5.20 shows the plot for site E1 for EFF-LAI using a CovMaterniso kernel with $V = 1$ summed with two Periodic kernels.
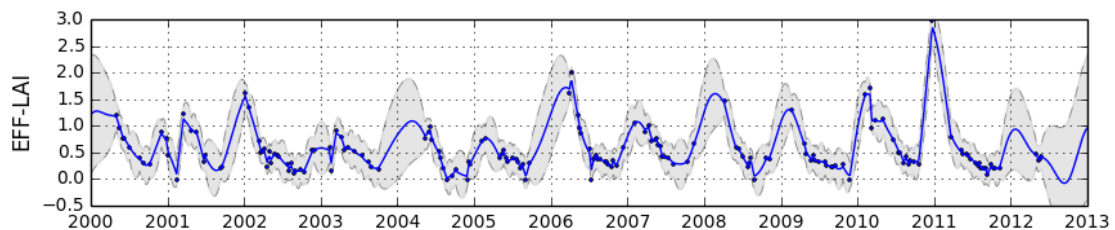

Figure 5.20: Site E1 - Alternative Kernels - EFF-LAI.

This combination of kernels does improve the NLL to a value of 12.948 (from 36.468 previously), although with noticeably increased roughness.

This raises the question of whether we are constrained to an ad-hoc strategy of simply trying different kernel combinations, until a suitable one is found. As noted in Chapter 2, for this dissertation, we have only adopted the approach of selecting common kernels "manually" and applying hyper-parameter optimisation as a method of modelling. In [Rasmussen and Williams 2006] however, more advanced methods are dealt with where the "model selection problem" is comprehensively addressed. The approach is to consider a higher-level of model which places a hierarchical Bayesian model over the hyper-parameters at the lower-level and a choice of covariance kernel functions or families at the higher-level. Both Bayesian and cross-validation methods of model-choice are described. These are aimed at more automated approaches to learning optimal models, but still within the Gaussian Process probabilistic machinery.

Taking cognisance of the circumspection that should be adopted when deriving interpretations from the RPV and TIP values in the absence of ancilliary data or proper validation studies (Chapter 2), we nevertheless put forward some conjectures on the patterns evident in the plots shown in this chapter.

Generally, phenology patterns can be seen in all of the sites, with the different variables showing different phases of phenology activity. For example, an increase in the RPV-K values in the dry season coinciding with a reduction in the FAPAR values. This could be associated with the background surface becoming more visible as the leaf-area index reduces, and hence the somewhat "dense" homogeneous view of the canopy from above becomes more "sparse" and 3-dimensional or heterogeneous.

There are also distinct differences in patterns across different years, but which are similar between sites, possibly giving information on more regional seasonal variability. The differences between sites

---

[3]The other neighbouring block, path 168/109 would also have the data, but was not provided as part of the set obtained from SANSA.

in particular seasons are more subtle, but evident. In particular, the fire event sites show noticeable differences especially in 2010 when comparing sites E3 and the fire sites F1 and F2. See for example the lower RPV-K values in 2010 compared to 2009 in Figure 5.16 for site F1 after the fire event during 2010 while site E3 (where there was no fire) maintains very similar values in 2009 and 2010.

When considering comparisons of site H1 bordering the park with the vegetation sites within the park, the plots presented are really only for illustration and no significant comments could be made about differences between the sites without further considerations. It would be interesting however, in possible future research, to determine whether the differences in the models could be interpreted quantitatively, or whether they simply fall within the natural variability of the system.

Although these observations have all been based on visual inspection of the plots, the curves generated could also be re-sampled to a desired density, and further analyses done computationally. These could include determining curve minima and maxima, gradient changes in sub-periodic regions, or generally applying the phenology activity detection methods cited in Chapter 2. This should also however, be qualified by stating that re-sampling is only making further predictions based on the original observations and any subsequent processing is inherently subject to the same uncertainties of the original model.

Figure 5.21 shows an example of alternating minimum and maximum values computed from the daily re-sampled predictions for RPV-K for site E1 (using the Composite Periodic kernel).
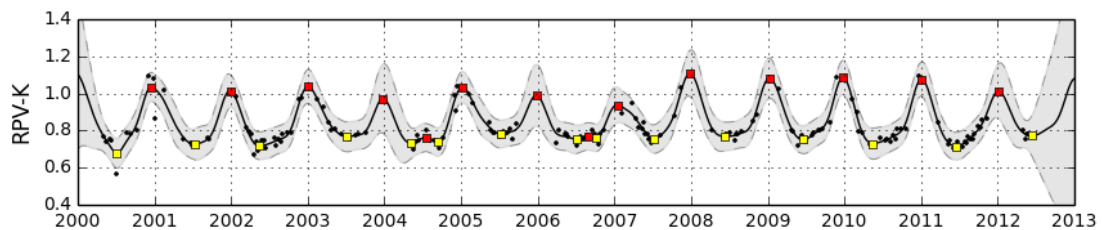

Figure 5.21: Site E1 - Minimum/Maximum - RPV-K.

Table 5.9 shows the dates and values of the RPV-K alternating minima and maxima.

Table 5.9: Site E1 Alternating Minimum/Maximum RPK-K Dates/Values

Date	Minimum	Date	Maximum
2000-07-05	0.678	2000-12-18	1.034
2001-07-11	0.726	2001-12-29	1.013
2002-05-15	0.717	2003-01-03	1.036
2003-07-07	0.766	2003-12-25	0.971
2004-05-07	0.731	2004-07-17	0.762
2004-09-12	0.738	2005-01-09	1.034
2005-07-09	0.781	2005-12-27	0.989
2006-07-05	0.752	2006-08-31	0.768
2006-10-03	0.760	2007-01-16	0.937
2007-07-10	0.754	2007-12-28	1.107
2008-06-12	0.767	2009-01-08	1.081
2009-06-18	0.751	2009-12-25	1.084
2010-05-17	0.729	2011-01-04	1.077
2011-06-15	0.715	2012-01-05	1.012
2012-06-10	0.772		

Figure 5.22 shows alternating minimum and maximum values computed from the daily re-sampled predictions for FAPAR for site E1 (using the Composite Periodic kernel).
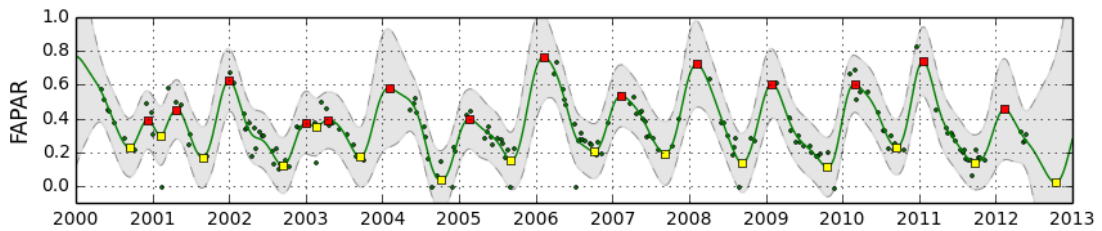

Figure 5.22: Site E1 - Minimum/Maximum - FAPAR.

Table 5.10 shows the dates and values of the FAPAR alternating minima and maxima.

Table 5.10: Site E1 Alternating Minimum/Maximum FAPAR Dates/Values

Date	Minimum	Date	Maximum
2000-09-14	0.226	2000-12-09	0.390
2001-02-09	0.298	2001-04-26	0.452
2001-08-27	0.168	2002-01-02	0.625
2002-09-11	0.121	2003-01-03	0.374
2003-02-20	0.349	2003-04-18	0.389
2003-09-17	0.171	2004-02-02	0.579
2004-10-06	0.034	2005-02-21	0.396
2005-09-04	0.149	2006-02-08	0.765
2006-10-03	0.205	2007-02-13	0.534
2007-09-10	0.192	2008-02-04	0.721
2008-09-11	0.135	2009-01-31	0.605
2009-10-20	0.112	2010-03-02	0.604
2010-09-17	0.230	2011-01-19	0.737
2011-09-23	0.139	2012-02-12	0.454
2012-10-12	0.019		

5.9   Conclusions

In this Chapter we have shown the process of using Gaussian Process methods for modelling data products from the MISR-HR system to describe the statistical properties of the time-series of the observations. The methods provide reasonable estimates of missing data and a basis for further investigating the vegetation phenology in the sample case areas.

Along with the single site evaluated in Chapter 4, the additional cases in this chapter would further support the assertion that Gaussian Process Regression is a viable method for time-series and temporal analysis.

As a postlude to this chapter, we show some further images of the areas in which the darkened patches are assumed to be the development of fire "scarring" around the sites chosen for the models. These are RGB images generated from the nadir camera BRF for the area over the period from September 2009 through to May 2011. Figures 5.23 and 5.24 show the development of the suspected fire patches in the centre of the images from 2009 to 2010, while Figure 5.25 shows that they are no longer evident by mid summer 2011 and have disappeared by the autumn (dry season) of 2011.
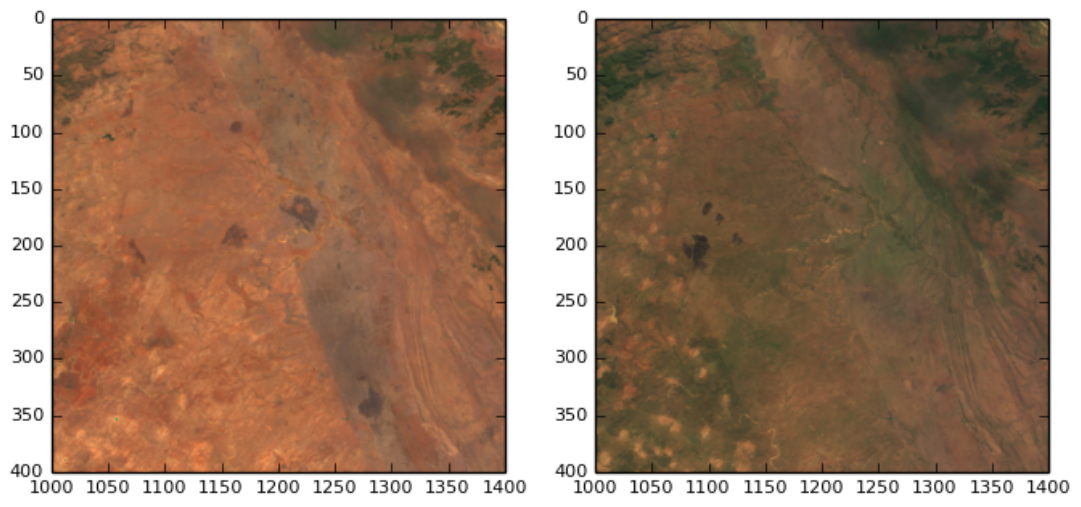
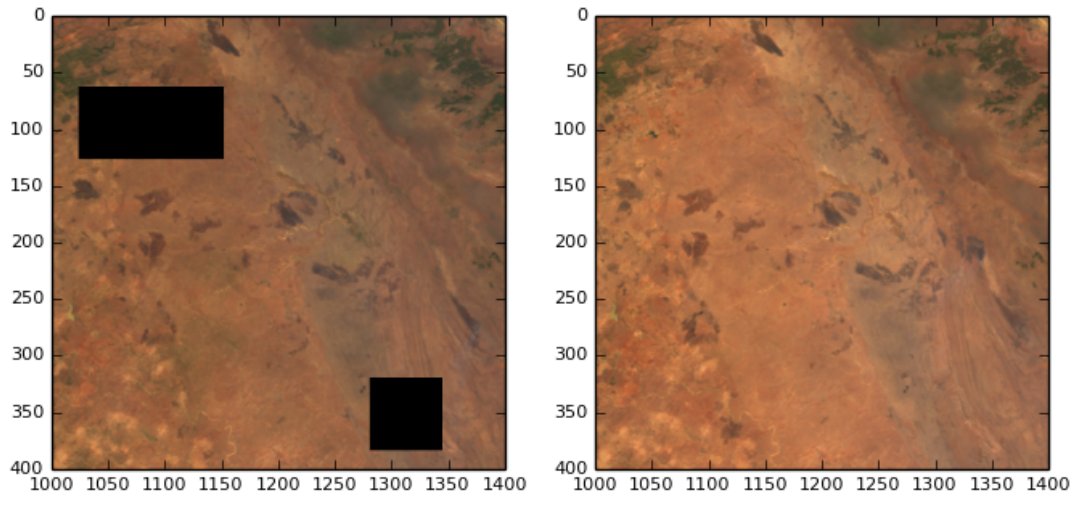Figure 5.23: Fire Activity 2009-09-16 and 2010-07-17.

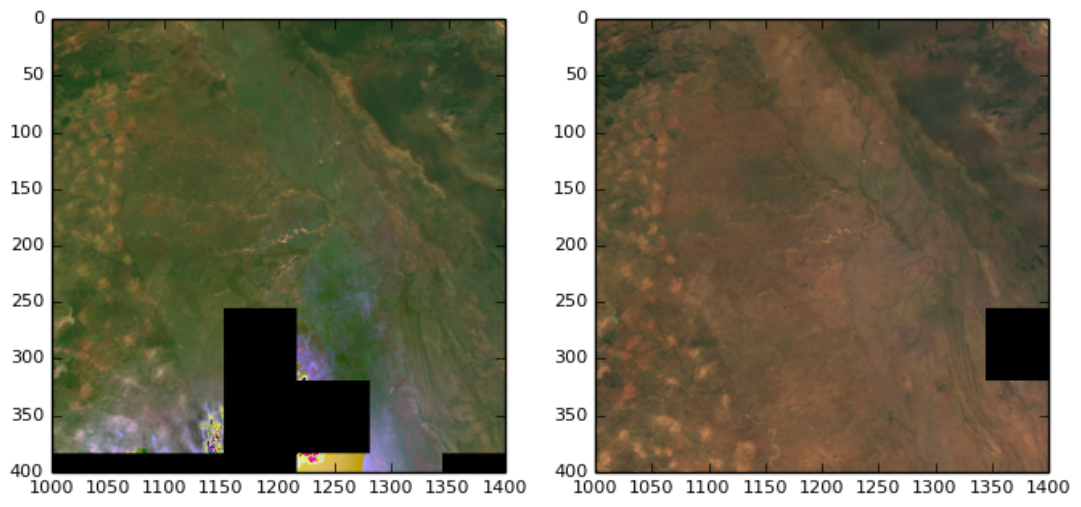Figure 5.24: Fire Activity 2010-09-19 and 2010-10-21.

Figure 5.25: Fire Activity 2010-12-24 and 2011-05-17.

Chapter 6

Conclusions

6.1 Summary

In this dissertation, we have undertaken an exploration of the MISR-HR system and the tremendously rich source of data and information that this system provides. Previous work over decades in evolving models such as RPV and TIP have provided the means of extracting highly meaningful and relevant information from satellite systems and the MISR satellite instrument in particular. The provision of this system and the data, for the African continent by NASA and the JRC to SANSA in South Africa, holds the potential for application of this to many needs in many domains in the country and across the continent.

To leverage the previous effective use of the Gaussian Procession Regression method in other remote-sensing domains, the aim of this research has been to assess whether this method could be viable for, and add value to, the present methods of extracting information from the MISR-HR and other similar satellite remote-sensing systems. The problem of resolving the characteristic irregularity and uncertainty in the availability of data for prescribed time-series analyses provided the research area in which Gaussian Processes could be considered. The context was constrained to time-series methods for applications of time-series in the analysis of vegetation phenology. With many existing methods available, the main motivation was to consider whether the Bayesian statistical roots of Gaussian Processes could be viable for this purpose, and bring the benefits of a single consistent framework in which the different requirements of time-series methods could be met. These requirements, established in previous work, formed the basis of the research questions to be assessed.

The methodology chosen was to undertake a practical study of applying the Gaussian Process methods to data from the MISR-HR system and to evaluate their viability in terms of these research questions. A case study for MISR-HR data provided by SANSA was undertaken for seven sites in the Kruger National Park region. The sites were chosen in which differing vegetation landscapes were known and historical data for a twelve year period was extracted and processed. A set of models for one site was developed comprehensively and assessed in detail against the research questions. Thereafter, the other sites were modeled using optimum model constructs from the first site. Generally Gaussian Process Regression was shown to support the assertion that it could be viable and that the single framework could deal with all of the requirements generally needed for time-series and phenology applications in satellite data.

A further contribution of this research was in undertaking the development in the Julia Scientific programming language, a recently available modern language platform. The software components developed for this project will be made available as Julia libraries or packages, once open source licensing

dependencies are ratified for the external libraries utilised.

6.2   Future Research

A further goal of this project had been to also consider the spatial aspects of the temporal patterns which could be modelled with Gaussian Processes. The Gaussian Process Regression method also provides for modelling regression in multiple dimensions. The temporal patterns from the different MISR-HR fields for the different sites showed that there are clear temporal "signatures" which could be used to distinguish or characterise spatially-temporal patterns as well. Further research along these lines was not possible in the context of the scope for this project, but could be interesting for future research.

Other areas for future research could be the investigation of more advanced Gaussian Process learning and optimisation methods to extend this work for MISR-HR. Many advanced alternatives have been and continue to be actively researched. Additionally, the complexities of rendering the complex information generated and modelled from the MISR-HR system for effective means of visualisation, would be valuable further research in this domain.

References

[Abrahamsen 1997] P. Abrahamsen. *A Review of Gaussian Random Fields and Correlation Functions. Technical Report 917, 2nd Ed.* Norwegian Computing Center, Oslo, Norway, 1997.

[Alpaydm 2010] E. Alpaydm. *Introduction to Machine Learning Second Edition*. The MIT Press, 2010.

[Atkinson *et al.* 2008] P. M. Atkinson, E. Pardo-Igu'zquiza, and M. Chica-Olmo. Downscaling cokriging for super-resolution mapping of continua in remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 2008.

[Atkinson *et al.* 2012] P. M. Atkinson, C. Jeganathan, J. Dash, and C. Atzberger. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123:400–417, 2012.

[Bezanson *et al.* 2012a] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman. Julia: A fast dynamic language for technical computing. 2012. arXiv:abs/1209.5145.

[Bezanson *et al.* 2012b] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman. *The Julia Language Platform*. http://www.julialang.org/, 2012. [accessed; 01-August-2013].

[Bishop 2006] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, 2006.

[Brahim-Belhouari and Bermak 2004] S. Brahim-Belhouari and A. Bermak. Gaussian process for non-stationary time-series prediction. *Computational Statistics & Data Analysis*, 47:705–712, 2004.

[Clerici *et al.* 2010] M. Clerici, M. Vossbeck, B. Pinty, T. Kaminski, M. Taberner, T. Lavergne, and I. Andredakis. Consolidating the two-stream inversion package (JRC-TIP) to retrieve land surface parameters from albedo products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3), 2010.

[Craven and Wahba 1979] P. Craven and G. Wahba. Smoothing noisy data with spline functions estimating the correct degree of smoothing by the method of generalized cross validation. *Numerische Mathematik*, 31:377–403, 1979.

[Cressie 1993] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1993.

[Diner *et al.* 1998] D. J. Diner, J. C. Beckert, T.H. Reilly, C.J. Bruegge, J.E. Conel, R.A. Hahn, J. V. Martonchik, T.P Ackerman, R. Davies, S. A. W. Gerstl, H. R. Gordon, J. P. Muller, R. B. Myeni, P. J. Sellers, B. Pinty, and M. M. Verstraete. Multi-angle imaging spectroradiometer (MISR) instrument description and experiment opverview. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4):1072–1087, 1998.

[Diner *et al.* 2002] D. J. Diner, M. M. Verstraete, and J. V. Martonchik. Foreword to special section on MISR. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7):1447–1448, 2002.

[Diner *et al.* 2007] D. J. Diner, L. Di Girolamo, and A. Nolin. Preface to the MISR special issue. *Remote Sensing Environment*, 107(1):1, 2007.

[Duvenard *et al.* 2013] D. Duvenard, J.R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. *JMLR: WCP, Proceedings of the 30th International Conference on Machine Learning*, 28, 2013.

[Fourier 1822] J. Fourier. *Theorie Analytique de la Chaleur*. Firmin Didot (reissued by Cambridge University Press), 1822.

[Gobron *et al.* 2002] N. Gobron, B. Pinty, M.M Verstraete, J. Widlowski, and D. J. Diner. Uniqueness of multiangular measurements - Part II: Joint retrieval of vegetation structure and photosynthetic activity from MISR. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7), 2002.

[Golyandina *et al.* 2001] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall, Washington DC USA, 2001.

[Grant and Thomas 2006] R. Grant and V. Thomas. *SAPPI Tree Spotting - Lowveld including The Kruger National Park*. Jacana Media (Pty) Ltd, 2006.

[Harris 1978] F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66:51–83, 1978.

[Hocke and Kämpfer 2009] K. Hocke and N. Kämpfer. Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle peridogram. *Atmosphere Chemistry and Physics*, 9:4197–4206, 2009.

[Hunter 2007] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[Hutchinson and de Hoog 1985] M. F. Hutchinson and F. R. de Hoog. Smoothing noisy data with spline functions. *Numerische Mathematik*, 47:99–106, 1985.

[Hutchinson 1988] M. F. Hutchinson. Interpolation of rainfall data with thin plate smoothing splines part i: Two dimensional smoothing of data with short range correlation. *Geographic Information and Decision Analysis*, 2(2):139–151, 1988.

[Johnson] Steven G. Johnson. http://ab-initio.mit.edu/nlopt. [accessed; 25-may-2013].

[Joubert 2007] S. Joubert. *The Kruger National Park*. High Branching (Pty) Ltd, 2007.

[Jovanovic *et al.* 2012] V. Jovanovic, K. Miller, B. Rheingans, and C. Moroney. *MISR Science Data Product Guide*. Jet Propulsion Labratory, California Institute of Technology, 2012.

[Koller and Friedman 2010] D. Koller and N. Friedman. *Probabilistic Graphical Models Principles and Techniques*. The MIT Press, 2010.

[Kondrashov and Ghil 2006] D. Kondrashov and M. Ghil. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Processes in Geophysics*, 13:151–159, 2006.

[Lauritzen 1981] S. L. Lauritzen. Time-series analysis in 1880, a discussion of contributions made by T. N. Thiele. *International Statistical Review*, 49:319333, 1981.

[Lavergne *et al.* 2007] T. Lavergne, T. Kaminski, B. Pinty, M. Taberner, N. Gobron, M. M. Verstraete, M. Vossbeck, J. Widlowski, and R. Giering. Application to MISR land products of an RPV Model Inversion Package using Adjoint and Hessian codes. *Remote Sensing of Environment*, 107:362–375, 2007.

[Liu and Nocedal 1989] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.

[Lomb 1976] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, 1976.

[Lubin and Dunning 2013] M. Lubin and I. Dunning. Computing in operations research using julia. *arXiv:1312.1431 [math.OC]*, 2013.

[MacKay 2003] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[Minnaert 1941] M. Minnaert. The reciprocity principle in lunar photometry. *Astrophysics Journal*, 93:403–410, 1941.

[Moffat *et al.* 2007] A. Moffat, D. Papale, M. Reichstein, D.Y. Hollinger, A. D. Richardson, A. G. Barr, C. Beckstein, B. H. Braswell, G. Churkina, A. R. Desai, E. Falge, J. H. Gove, M. Heimann, D. Hui, A. J. Jarvis, J. Kattge, A. Noormets, and V.J. Stauch. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forrest Meteorology*, 147:209–232, 2007.

[Musial *et al.* 2011] J. P. Musial, M.M Verstraete, and N. Gobron. Comparing the effectiveness of recent algoritms to fill and smooth incomplete and noisy time series. *Atmosphere Chemistry and Physics*, 11:7905–7923, 2011.

[Nickisch and Rasmussen 2008] H. Nickisch and C. E. Rasmussen. Approximations for binary gaussian process classification. *Journal of Machine Learning Research*, 9:2035–2078, 2008.

[Nocedal 1980] J. Nocedal. Updating quasi-newton matrices with limited storage. *Mathematical Computation*, 35:773–782, 1980.

[Osborne *et al.* 2008] M. A. Osborne, A. Rogers, S Ramchurn, S. J. Roberts, and N. R. Jennings. Towards real-time information processing of sensor network data using computationally efficient multi-output gaussian processes. *International Conference on Information Processing in Sensor Networks*, pages 109–120, 2008.

[Pinty *et al.* 2002] B. Pinty, J. Widlowski, N. Gobron, M.M Verstraete, and D. J. Diner. Uniqueness of multiangular measurements - Part I: An indicator of subpixel surface heterogeneity from MISR. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7), 2002.

[Powell 1994] M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. *Advances in Optimization and Numerical Analysis, eds. S. Gomez and J.-P. Hennart (Kluwer Academic: Dordrecht)*, pages 51–67, 1994.

[Powell 1998] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[Press *et al.* 1992] W. H. Press, S. A. Teukolsky, W.T. Vetterling, and B. P. Flannery. *Numerical recipes in Fortran, 2nd Ed*. Cambridge University Press, Cambridge, USA, 1992.

[Rahman *et al.* 1993a] H. Rahman, B. Pinty, and M.M Verstraete. Coupled Surface-Atmosphere Reflectance (CSAR) Model 1. model description and inversion against synthetic data. *Journal of Geophysical Research*, 98(D11):20779–20879, 1993.

[Rahman *et al.* 1993b] H. Rahman, B. Pinty, and M.M Verstraete. Coupled Surface-Atmosphere Reflectance (CSAR) Model 2. semiempirical surface model usable with NOAA Advanced Very High Resolution Radiometer Data. *Journal of Geophysical Research*, 98(D11):20791–20801, 1993.

[Rasmussen and Nickisch 2010] C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (gpml) toolbox. *Journal of Machine Learning Research*, 11:3011–3015, 2010.

[Rasmussen and Williams 2006] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[Reinsch 1967] C. M. Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10:177–183, 1967.

[Roberts *et al.* 2012] S. Roberts, A. Rogers, N. Jennings, and M. Osborne. Real-time information processing of environmental sensor network data. *Transactions on Sensor Networks*, 9(1), 2012.

[Roberts *et al.* 2013] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 371(1984), 2013.

[Services 2014] NASA Langley ASDC User Services. *MISR - Multi-angle Imaging SpectroRadiometer*. http://www-misr.jpl.nasa.gov/, 2014. [accessed; 08-january-2014].

[The Gaussian Process Web Site] http://www.gaussianprocess.org. [accessed; 24-April-2013].

[The HDFEOS Libraries Web Site] http://hdfeos.org/software/library.php. [accessed; 26-February-2013].

[Venter and Gertenbach 1986] F. J. Venter and W. P. D. Gertenbach. A cursory review of the climate and vegetation of the kruger national park. *Koedoe*, 29:139–148, 1986.

[Verstraete *et al.* 1990] M.M. Verstraete, B. Pinty, and R.E. Dickinson. A physical model of the bidirectional reflectance of vegetation canopies, 1, theory. *Journal of Geophysical Research*, 95:11755–11765, 1990.

[Verstraete *et al.* 2007] M. M. Verstraete, N. Gobron, O. Aussedat, M. Robustelli, B. Pinty, J. Widlowski, and M. Tabener. An automatic procedure to identify key vegetation phenology events using the JRC-FAPAR products. *Journal of Advances in Space Research*, pages 1773–1783, 2007.

[Verstraete *et al.* 2012] M. M. Verstraete, L. A. Hunt, R. J. Scholes, M. Clerici, B. Pinty, and D. L. Nelson. Generating 275-m resolution land surface products from the Multi-Angle Imaging SpectroRadiometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10):3980–3990, 2012.

[Whittaker 1923] E. T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1923.

[Xie *et al.* 2008] Y. Xie, Z. Sha, and M. Yu. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1(1):9–23, 2008.