

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
Maths 2018; 3(4): 109-120  
© 2018 Stats & Maths  
www.mathsjournal.com  
Received: 11-05-2018  
Accepted: 12-06-2018

## E Okuto

School of Mathematics &  
Actuarial Science, Jaramogi  
Oginga Odinga University of  
Science & Technology, Bondo-  
Usenge Road, P.O Box 210-  
40601, Bondo, Kenya

## B Omolo

Division of Mathematics and  
Computer Science, University of  
South Carolina - Upstate, 800  
University way, Spartanburg,  
SC, 29303, USA

## O Ongati

School of Mathematics &  
Actuarial Science, Jaramogi  
Oginga Odinga University of  
Science & Technology, Bondo-  
Usenge Road, P.O Box 210-  
40601, Bondo, Kenya

## Correspondence

### E Okuto

School of Mathematics &  
Actuarial Science, Jaramogi  
Oginga Odinga University of  
Science & Technology, Bondo-  
Usenge Road, P.O Box 210-  
40601, Bondo, Kenya

## Reconstructing global earth observation based vegetation index records with stochastic partial differential equations approach

E Okuto, B Omolo and O Ongati

### Abstract

Long-term Earth observation based vegetation index records have been used extensively by researchers to assess vegetation response to global climate variability and change. However, the records exhibit multiple temporal gaps due to spectral and radiometric inconsistencies that inhibit accurate assessment of land surface vegetation dynamics. Here, we propose a new reconstruction procedure that approximates Bayesian time series model by using integrated nested Laplace approximations (INLA) to overcome Bayesian computational limitations. The technique was tested on the vegetation index and phenology (VIP) Lab enhanced vegetation index-two (VIP-EVI2) version 3 15-day 5 km resolution record. VIP-EVI2 is a reconstructed record with inverse distance weighting function and linear interpolation (IDW-EVI2). VIP-EVI2 is derived from red and near-infrared (NIR) top of canopy (TOC) reflectance, detected by the Advanced Very High Resolution Radiometer (AVHRR). The INLA-EVI2 was compared globally and locally with an adaptive Savitzky-Golay (SG-EVI2) filter. The global evaluation was done by descriptive analysis, goodness-of-fit by Kolmogorov-Smirnov (K-S) test, annual trend analysis by Thiel-Sen (T-S) slope. The local comparison was done by evaluating the ability of IDW-EVI2, SG-EVI2, and INLA-EVI2 to estimate *in situ* Leaf Area Index (LAI) measurements taken over several years and for major field crops across the globe. Locally, INLA-EVI2 estimated the *in situ* data more correctly than SG-EVI2 as indicated by  $R^2$  and RMSE. Globally, the INLA-EVI2 recorded a better goodness-of-fit, more stable and consistent trends than SG-EVI2. Based on these findings, if computational resources are unlimited, the INLA approach provides a viable alternative to standard reconstruction procedures.

**Keywords:** Advanced very high resolution radiometer (AVHRR), vegetation index and phenology (VIP) lab, enhanced vegetation index (EVI), Gap-filling, bayesian inference, integrated nested laplace approximation (INLA)

### 1. Introduction

The long-term monitoring of global processes and feedbacks is important for assessing the impact of climate variability and change on ecosystems (Batista *et al*, 1997; Weiss *et al*, 2004) [39]. Global vegetation records over the past three decades have facilitated studies dealing with global carbon flux, land use/cover change, and crop production estimation (Olsson *et al*, 2005; Xia *et al*, 2008) [26, 40]. These vegetation records generally consist of the Normalized Difference Vegetation Index (NDVI) derived from top of canopy (TOC) visible red and near-infrared (NIR) reflectance. In general, reflectance is detected by the Advanced Very High Resolution Radiometer (AVHRR) pre 2000, which is carried onboard the National Oceanic and Atmospheric Administration (NOAA) polar orbiting operational environmental satellites and moderate imaging spectroradiometer (MODIS) post 2000, aboard NASA's Aqua satellite (Barreto-Munoz, 2013) [2].

AVHRR is available for 30+ years, much suited for long-term climate change research studies that require going back to the 80's and 90's for more accurate global change detection (Brown *et al.*, 2006; Marshall *et al.*, 2016) [4, 23]. However, for new studies (2000+), moderate imaging spectroradiometer (MODIS) (Fang *et al.*, 2014; Lu *et al.*, 2015) [7, 22] is more recommended given its higher spatial, spectral resolution, and ability to eliminate background and atmosphere noises. NDVI is developed from a ratio of NIR minus visible red to NIR plus visible red to NIR plus visible red. Vegetation cover scatters (absorbs) strongly in the NIR (red) producing high NDVI compared to bare soil, which scatters strongly in both the NIR and

red (Peñuelas and Filella, 1998) <sup>[28]</sup>. Aerosols, dust, Rayleigh scattering and partial cloud cover tend to increase reflectance in the NIR and lower NDVI, thus requiring additional compositing before interpretation (Holben, 1986) <sup>[14]</sup>. Major cloud contamination, particularly in the tropics, can create numerous gaps in the daily AVHRR records, eliminating approximately two-thirds of the data (Justice *et al.*, 1991) <sup>[18]</sup>. One way to minimize the sensor and atmospheric effects is by using 15-day maximum value composites (MVC), however, gaps and inconsistencies persist, often necessitating further gap-filling and smoothing.

Several long-term vegetation records exist, each with different approaches to standardize NDVI across the NOAA satellites and to account for atmospheric contamination and other effects. The Global Inventory Monitoring and Mapping Studies (GIMMS) is the most widely used NDVI product and provides long-term global data at 8 km spatial resolution (Fensholt *et al.*, 2006) <sup>[8]</sup>. GIMMS adopts an Empirical Bayesian technique to statistically correct parts of the vegetation record that could be affected by Satellite drift and atmospheric contamination (Pinzon and Tucker, 2014; Tucker *et al.*, 2005) <sup>[29, 36]</sup>. NDVI tends to saturate more in dense vegetation and is more sensitive to soil background than enhanced vegetation index (EVI), hindering proper estimation of important biophysical parameters such as vegetation fraction and leaf area index (LAI) (Huete *et al.*, 2002; Zhao *et al.*, 2012) <sup>[17, 24]</sup>, so a new long-term EVI product has been developed by the Vegetation Index & Phenology (VIP) Lab (Didan, 2014) <sup>[6]</sup> using AVHRR and MODIS data to overcome the limitations of NDVI. Unlike NDVI, EVI does not saturate severely in dense vegetation and incorporates a correction term to reduce the effects of soil background. The standard EVI is computed with visible red, NIR, and blue reflectance from MODIS. Since blue reflectance is not available from AVHRR, VIP developed a visible red-NIR version of EVI called EVI2 (Jiang *et al.*, 2008) <sup>[17]</sup>.

Unlike GIMMS, the VIP product which is essentially a fusion of AVHRR and MODIS undergoes partial atmospheric correction at a daily time step, before it is compositing. VIP-EVI2 and MODIS-EVI are highly correlated and perform better than their NDVI counterpart when predicting ground-based LAI (Rocha and Shaver, 2009) <sup>[31]</sup>. Due to the processing steps, however, VIP is less consistent temporally than GIMMS resulting to poor performance in trend analysis (Tian *et al.*, 2015) <sup>[35]</sup>, and has a higher bias with ground data as noted in Marshall *et al.* (2016) <sup>[20]</sup>. This is due in part to the way VIP-EVI2 was reconstructed using inverse distance weighting (IDW) and linear interpolation that tended to suffer from lack of representation of variability associated with the averaged values over various regions (Zhang *et al.*, 2014) <sup>[40]</sup>. Given that the effectiveness of compositing is limited, it has been suggested that representing bad-quality records as expressed by quality information with gaps followed by a reconstruction could improve the continuity and quality of the vegetation index data (Kandasamy *et al.*, 2013; Pinzon & Tucker, 2014; Zhou *et al.*, 2015) <sup>[19, 29, 42]</sup>.

Adaptive Savitzky-Golay (SG), Whittaker-Henderson (WH) and IDW are some of the most commonly used reconstruction techniques in remote sensing. Conceptually, SG is a piece-wise regression procedure while WH is based on the minimization of a cost function describing the balance between fidelity and roughness. While SG and WH are curve-fitting techniques, IDW is a deterministic spatial interpolation technique in which values assigned to unknown points are calculated with an average, weighted by the inverse of the distance to each neighboring known point (Daly *et al.*, 2002) <sup>[5]</sup>. As noted by Kandasamy *et al.* (2013) <sup>[19]</sup>, accuracy of SG and WH are significantly affected by the percentage of missing observations before reconstruction. Specifically, SG requires an *a priori* smoothing parameter and window size, which are assumed to be fixed while for WH, a user specified smoothing parameter is necessary to control the balance between fidelity and roughness. Vegetation index records are dynamic and as such it would be expected that multiple parameters and/or hyper-parameters would be required to characterize its variability. Consequently, a model that is driven by one (e.g. WH) or two (e.g. SG) parameters may not account for the full changes in location, variability, and shape that do occur in physical space and time (Weiss *et al.*, 2014) <sup>[38]</sup>. An adaptive SG routine that allows the smoothing parameters to vary according to optimization criteria has been suggested by Chen *et al.* (2004) <sup>[7]</sup>. Further improvements have been suggested to overcome the challenge of over-smoothing or persistent roughness, but tend to perform better with pixel-level quality information characterized with large persistent good quality data. Full Bayesian inference is very flexible and does not require subjective user input and is particularly recommended for persistent large gaps with potential outliers in a long-term time series vegetation index record (Katzfuss and Cressie, 2011) <sup>[21]</sup>. Unfortunately, full Bayesian inference is computationally more intensive than other techniques and has therefore not been widely used (Pinzon and Tucker, 2014) <sup>[29]</sup>. Maximum likelihood techniques have been successfully adopted to identify optimal parameters to reduce computational constraints (e.g. empirical Bayes) improving feasibility (Pinzon and Tucker, 2014) <sup>[29]</sup>.

This study considers an approximate Bayesian inferential procedure designed for latent Gaussian models to reconstruct long-term Earth observation records that is flexible and efficient (Rue *et al.*, 2009) <sup>[32]</sup>. The Markov chain Monte-Carlo (MCMC) model convergence limitation, which accounts for most of computational demand, was overcome using an Integrated Nested Laplace Approximations (INLA) algorithm that utilizes sparse precision matrices. Physical time series components (trend, seasonal, and cyclical) are adjusted in an additive way using latent time series functions and pixel-level predictive distributions extracted as the unbiased estimate of the missing values.

The paper is organized as follows: In section 2.1, EVI2 is described; in sections 2.2 and 2.3, a brief theoretical background of the Bayesian time series model using INLA is presented, and the adaptive SG routine for gap-filling and smoothing Earth observation data for comparison is briefly presented in section 2.4. An overview of the analytical methods used to compare the techniques is provided in sections 2.5 and 2.6. Results of the inter-comparison are provided in section 3.1 and 3.2, followed by a discussion of the results and conclusion in sections 4 and 5, respectively.

## 2. Data, Processing and Methods

### 2.1 Vegetation Index & Phenology (VIP) Lab Version 3 Enhanced Vegetation Index 2 (EVI2)

We used the VIP-EVI2 record for comparison, given its potential advantage over NDVI for biophysical modeling and persistent data gaps that remain in the record after pre-processing. The data can be downloaded at [http://vip.arizona.edu/viplab\\_data\\_explorer.php](http://vip.arizona.edu/viplab_data_explorer.php). VIP-EVI2 is a multi-sensor product that is partially atmospherically corrected and is available globally at a 7-day, 15-day and monthly time steps from 1981 to present at 0.05° (~5.6 km at the equator) spatial resolution (Jiang *et al.*, 2008) <sup>[17]</sup>. For purposes of this study, 15-day VIP-EVI2 version 3 data from 1982-2011 was used. At the

time of reconstruction, data analysis and writing the manuscript, a newer version of VIP-EVI2 became available, but since the purpose of this manuscript was to demonstrate the improved accuracy and lower computational demands of a Bayesian smoothing and gap-filling technique, and not the dataset itself, we retained version 3.0. Like MODIS, VIP-EVI2 includes pixel reliability (PR) bands that characterize the surface reflectance state, condition of the atmosphere and other useful information about each pixel (Didan, 2014) [6]. PR is indexed from 0 to 7 denoting best to worst data quality respectively. In its native form, Inverse Distance Weighting and Linear Interpolation (IDW) are used to reconstruct the record in order to reflect a continuous phenology curve (Didan, 2014) [6]. However, in this study, pixel level artificial gaps were created for data with corresponding PR>1, because such pixels were significantly affected by clouds, aerosols, dust, Rayleigh scattering, and other noise. This resulted in large, persistent and irregularly spaced data gaps beyond the ability of the smoothing and gap-filling methods evaluated in this study. So, these gaps were partially filled before the assessment using the MODIS filtering algorithm proposed in Xiao *et al.* (2003) [40] ; Marshall *et al.* (2016) [20] and adopted by Opiyo *et al.* (2013) for the tropics. The resulting pixel level percentage records available after filtering, but before applying gap-filling and smoothing techniques is provided in Figure 1. For the purpose of this study, the native gap-free record will be referred to as IDW-EVI2 while the record with missing bad-quality data referred to as VIP-EVI2.

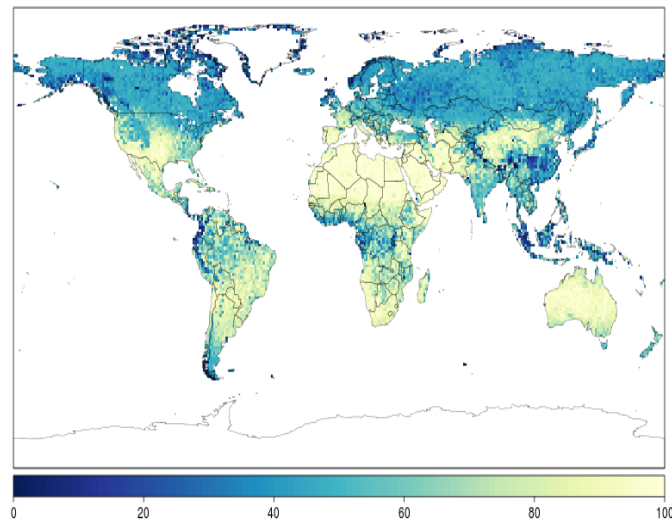


Fig 1: The percentage of data available after filtering, but before applying gap-filling and smoothing.

Further details on the dataset can be found in Rocha and Shaver (2009) [31]. EVI2 is derived from the following equation

$$EVI2 = 2.5 \frac{NIR-RED}{NIR+2.4 RED+1} \tag{1}$$

### 2.2 Reconstruction with Bayesian time series model using INLA

Vegetation index records can be classified as a hierarchical Gaussian process with linear predictor

$$\eta_j = \beta_0 + \sum_{k=1}^{n\beta} \beta_k z_{ki} + \sum_{k=1}^{nf} f^{(i)}(u_{ij}) + \epsilon_i, \tag{2}$$

Where  $\{\beta_k\}$ 's represents the linear effects of covariates  $z$  (e.g. temperature, precipitation etc.),  $\{f^{(i)}(\cdot)\}$ 's are non-linear functions of the covariates  $u$  (e.g. trend, seasonal, and cyclic effects) used to relax the linearity of the process, and  $\epsilon_i$ 's are unstructured terms. This type of model is considered as a special case of the more general class of latent Gaussian models for which approximate Bayesian inference can be performed using Integrated Nested Laplace Approximations (INLA).

Latent Gaussian models are considered as a sub class of structured additive regression models, in which the predictor can be expressed as a function of linear and nonlinear effects of the covariates. The vegetation index record  $y_j$  in a given pixel is assumed to belong to an exponential family where the mean  $\mu_j$  is linked to a structured additive predictor  $\eta_j$  through a link function  $g(\mu_j) = \eta_j$ .

In this study, trend, seasonal, and cyclic terms are added to relax the linearity of the long-term time series vegetation index record. For instance, physical long-term trend effect is developed from a sequence between one and the record length. In addition, seasonal annual (short-term) variability is created and adjusted for by replicating between 1 and 12 each twice (15a and 15b) repeating the resulting sequence number of study years times. Also, long-term cyclic component adjusting for potential climate variability and change effect is developed from a replication between one and number of study years each repeated number of annual records times. The model formula is then expressed as follows;

$$g(\mu_j) = \eta_j = \beta_0 + f(z_1, \text{model} = \text{ar1}) + f(z_2, \text{model} = \text{season}, \text{season.length} = n.\text{season}) + f(z_3, \text{model} = \text{season}, \text{season.length} = n.\text{cyclic}) \tag{3}$$

Where  $\beta_0$  is the model intercept,  $z_1, z_2, z_3$  are the trend, season, and cyclic covariates respectively,  $n.\text{season}$  and  $n.\text{cyclic}$  are to be determined based on the short term and long term seasonal patterns of the process which for this study were specified as 12 and 30 respectively while the model season is as provided by default with INLA.

An autoregressive model of order 1 ( $ar1$ ) for the Gaussian process  $y(y_1, \dots, y_{720})$  is defined as  $y_1 \sim N(0, (\tau(1 - \rho^2))^{-1})$  and autocorrelation between neighboring outcomes say  $y_{i-1}$ ,  $y_i$ , and  $y_{i+1}$  is expressed such that

$$y_i = \rho y_{i-1} + \epsilon_i \quad (4)$$

Where  $\epsilon_i \sim N(0, \tau^{-1})$   $i = 2, \dots, n$  for  $|\rho| < 1$ . The model for seasonal variation for periodicity say  $m$  for the random vector  $y(y_1, \dots, y_{720})$ ,  $n > m$  is obtained assuming that the sums  $y_i + y_{i+1} + \dots + y_{i+m-1}$  are an independent Gaussian process with precision  $\tau$  and the density for  $y$  is derived from the  $i + m - 1$  increments as;

$$\pi(y|\tau) = \tau^{\binom{n-m+1}{2}} \exp\left\{-\frac{1}{2} y' Q y\right\} \quad (5)$$

Where  $Q = \tau R$  and  $R$  is the structure matrix reflecting the neighborhood structure of the model.

The terms are collected in a latent field  $\Psi = \{\beta_0, \beta, f\}$  and assigned Gaussian priors such that the resulting model can be viewed as a latent Gaussian model. The posterior marginal of each element  $\psi_j$  of the latent field can be expressed by;

$$\pi(\psi_j|y) = \int \pi(\psi_j|\theta, y) \pi(\theta|y) d\theta \quad (6)$$

Where, the vector  $\theta$  denotes the hyper parameters of the model including those defining prior distributions. Applying the INLA methodology, the marginals are estimated by combining an analytical approximation with numerical integration (Rue *et al.*, 2009) [32]. The posterior distribution for bad-quality data can be obtained by integrating out Equation 7

$$\pi(y_j|y_{-j}, \theta) = \int \pi(y_j|\psi_j, \theta) \pi(\psi_j|y_{-j}, \theta) d\psi_j \quad (7)$$

The mean of the posterior distribution  $\{\pi(y_j|y_{-j}, \theta)\}$  is extracted and used as the unbiased estimator to the missing/bad-quality data. The Bayesian time series procedure will be referred to as INLA method while reconstructed record using the technique referred to as INLA-EVI2.

### 2.3 Adaptive Savitzky-Golay Filtering technique

Performance evaluation of the Bayesian time series model was compared to the adaptive Savitzky-Golay (SG) routine proposed by The adaptive SG has been widely used to fill MODIS data. The function was implemented using the *MODIS* package in R. As in other reconstructions, padding was added at the beginning and end of the time series, so that smoothing and gap-filling could be performed on the tails. For the remainder of this paper, the reconstructed record using the SG procedure will be referred to as SG-EVI2.

### 2.4 Gap-filling and smoothing technique (global inter-comparison)

INLA-EVI2 was compared to the SG-EVI2 and IDW-EVI2 both globally and locally. The global inter-comparison was made by assessing whether IDW-EVI2 (reference record), INLA-EVI2 and SG-EVI2 were from an identical statistical distribution. The magnitude of departure from equality was measured using the distance statistic (D) two sample Kolmogorov-Smirnov (K-S) test. In addition, the ability of the SG-EVI2, INLA-EVI2, and IDW-EVI2 to produce consistent records was evaluated using long-term trend analysis based on a non-parametric Thiel-Sen (T-S) regression (Fensholt and Proud, 2012) [9] masked for significance at  $\alpha=0.05$  estimated using a seasonally corrected Mann-Kendall trend test.

#### 2.4.1 Goodness-of-fit test with two-sample Kolmogorov-Smirnov test

The two-sample Kolmogorov-Smirnov (K-S) test was employed to assess degree of similarity between IDW-EVI2 (reference record), SG-EVI2 and INLA-EVI2 time series records. It is based on a statistical distance statistic (D) procedure involving two sample records, with a low value of D statistic ( $p\text{-value} > 0.05$ ), suggesting increased accuracy of the method. High value of D ( $D > 0.5$ ) tended to produce low  $p\text{-values}$  ( $p\text{-value} < 0.05$ ) suggesting varied statistical distribution. The non-parametric technique is a widely used statistical procedure that compares two empirical continuous distribution functions. The D statistic is computed from the maximum distance between the two empirical distributions based on the ranks. Although the test analyzes the actual data, it is equivalent to analysis of ranks which tend to be robust and desirable when the underlying statistical distribution of the records is not well understood.

#### 2.4.2 Long-term Trend Analysis with Theil-Sen Regression

The ability of the smoothing techniques to produce consistent long-term time series records was evaluated using the Theil-Sen (T-S) regression approach. It computes pairwise slopes of all datapoints and computes the median slope of those values. The median slope value becomes unbiased slope estimate.

#### 2.4.3 Mann-Kendall Slope Significance Testing

Statistical significance of the T-S regression slope estimates was evaluated using the Mann-Kendall (M-K) test. Slope significance with M-K test was considered due to its flexibility when the record is suspected to be non-linear with potential outliers. It is a non-parametric distribution free test for monotonic trends and correlation analysis. The test statistic is then provided by the number of positive differences minus the number of negative differences (Verbesselt *et al.*, 2010) [35]. A positive test statistic implies that observations obtained later in time tend to be larger than observations made earlier, while a negative test statistic implies that the

observations made later tend to be smaller than those made earlier (Tian *et al.*, 2015) [35]. For every test statistic, a probability value was extracted and used to mask T-S slope estimates for significance at  $\alpha=0.05$ .

## 2.5 Gap-filling and smoothing technique (local inter-comparison)

The new smoother was compared to the optimization routine locally (subset of pixels) using a newly developed database of 1,459 *in situ* LAI measurements for major field crops (alfalfa, barley, canola, cotton, garlic, maize, onion, pasture, potato, rice, soybean, sugar beet, and wheat) on five continents (Asia, Australia, Europe, North America, and South America). The database was compiled from several sources of LAI measurements made using destructive and non-destructive (optical) methods spanning the evaluation period. NDVI and other vegetation indices are widely used to compute LAI in global change studies (Fensholt *et al.*, 2004; Gitelson *et al.*, 2014; Potitthep *et al.*, 2010; Wang *et al.*, 2005) [10, 12, 30, 22], so Landsat reflectance data (30 m resolution) was later processed to compute NDVI and other vegetation indices, and paired with the *in situ* LAI data to evaluate the universality of the vegetation index-LAI relationship. Further details on this dataset can be found in Kang *et al.* (2016) [20, 23].

Since the relationship between *in situ* and coarse resolution vegetation indices and LAI is typically non-linear (Asrar *et al.*, 1992; Myneni *et al.*, 2002; Friedl *et al.*, 1995; Gutman and Ignatov, 1998; Sellers, 1985) [1, 25, 11, 13], Landsat EVI2 pixels within each IDW-EVI2 pixel were used to downscale IDW-EVI2, SG-EVI2, and INLA-EVI2 data using the vegetation fraction ( $F_C$ ) procedure suggested by Hwang *et al.* (2011) [15]. Unlike the vegetation index-LAI relationship, the vegetation index -  $F_C$  relationship is quasi-linear; meaning  $F_C$  estimated with a coarse resolution pixel is approximately equal to the average of  $F_C$  estimating from corresponding higher resolution pixels. In Hwang *et al.* (2011) [15], this relationship was used to downscale MODIS (250 m) NDVI with Landsat NDVI. The technique is ratio-based and parameterized for flat and complex terrain. Since the LAI measurements were retrieved from large and flat agroecosystems, we used the simple proportionality formula for each data pair to convert IDW-EVI2 to downscaled (30 m)  $F_C$  for comparison with *in situ* LAI:

$$\alpha = F_{C, \text{Landsat}} / \text{EVI2}_{\text{Landsat}} \quad (8)$$

Where  $\alpha$  is the proportionality constant,  $F_{C, \text{Landsat}}$  is the vegetation fraction computed from Landsat EVI2, and  $\text{EVI2}_{\text{Landsat}}$  is EVI2 derived from Landsat top of atmosphere reflectance. We computed  $F_C$  for Landsat with the formula developed for the MODIS Evapotranspiration Model (MOD16)

$$F_{C, \text{Landsat}} = (\text{EVI2} - \text{EVI2}_{\min}) / (\text{EVI2}_{\max} - \text{EVI2}_{\min}) \quad (9)$$

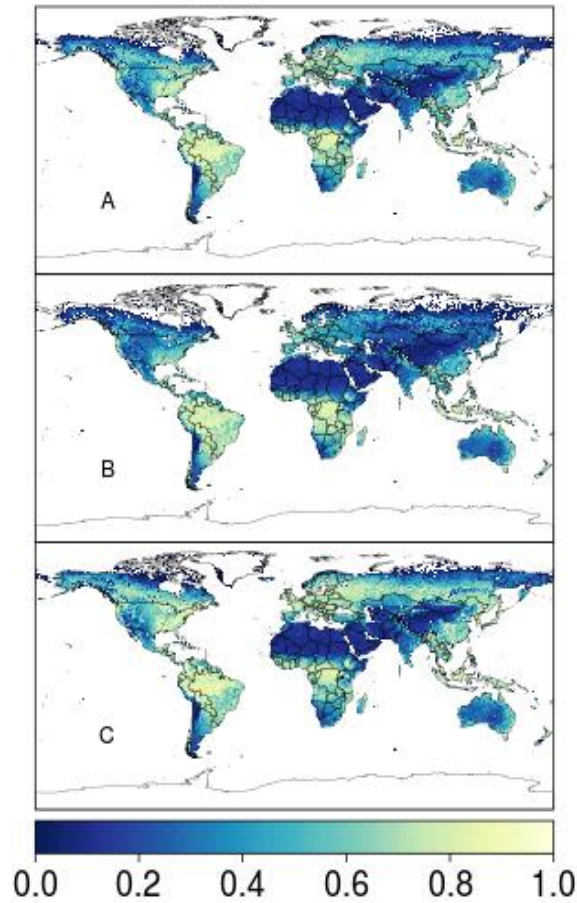
Where  $\text{EVI2}_{\min}$  equals 0.05 and represents EVI2 for bare soil ( $\text{LAI} \rightarrow 0$ ) and  $\text{EVI2}_{\max}$  equals 0.95 and represents EVI2 for dense vegetation ( $\text{LAI} \rightarrow \infty$ ). IDW-EVI2  $F_C$  was computed by multiplying IDW-EVI2 by the proportionality constant. In some cases, more than one Landsat pixel fell within an IDW-EVI2 pixel in space and time. In these cases,  $F_{C, \text{Landsat}}$  values were averaged before downscaling. In order to compare SG-EVI2 and INLA-EVI2, Landsat-LAI pairs were removed and then the corresponding time series were smoothed and gap-filled. Standard linear regression statistics (coefficient of determination- $R^2$  and Root Mean Squared Error-RMSE) were computed from the raw IDW-EVI2 data, SG-EVI2, and INLA-EVI2 versus LAI. Transformations were deemed unnecessary, because the data exhibited near linear relationships (not shown).

## 3. Results

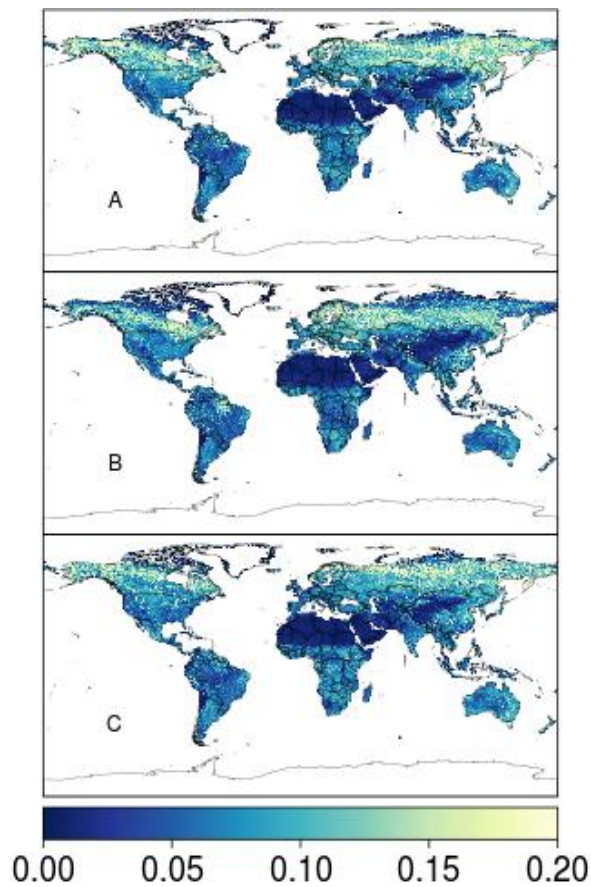
### 3.1 Global inter-comparison

Pixel level means and standard deviations were obtained to compare the distribution of SG-EVI2 and INLA-EVI2 to IDW-EVI2 (Figure 2). SG-EVI2 tended to over-predict IDW-EVI2, while INLA-EVI2 tended to under-predict IDW-EVI2 as shown in Figure 2. Although the absolute magnitudes were different, INLA-EVI2 showed a higher level of agreement with IDW-EVI2 in terms of the overall distribution of estimated values than SG-EVI2. The greatest difference between the SG-EVI2 and INLA-EVI2 was at the extreme Northern latitude in which the mean predicted values appear lowest and most variable (Figure 3).

The two sample K-S test was used to assess whether INLA-EVI2 and SG-EVI2 had identical distributions to IDW-EVI2 (in terms of medians, variability, and shape). The null hypothesis is that the IDW-EVI2 and SG-EVI2 or INLA-EVI2 was sampled from populations with identical distributions. The magnitude of the D statistic expresses indirectly, the degree of violation of the null hypothesis with higher values ( $p < 0.05$ ) suggesting that the two reconstructed records are from different statistical distributions. Overall goodness-of-fit on a monthly basis for SG-EVI2 and INLA-EVI2 are shown in Figure 4 and Figure 5, respectively. In each case, pixel level K-S estimates with  $p$ -value  $< 0.05$  were masked for display purposes. As shown in Figures 4 and 5, the patterns in correlations between IDW-EVI2 (reference), SG-EVI2, and INLA-EVI2 based on the D statistic were fairly identical. Nevertheless, INLA-EVI2 and IDW-EVI2 appears to be more similar as suggested by lowest D statistic.

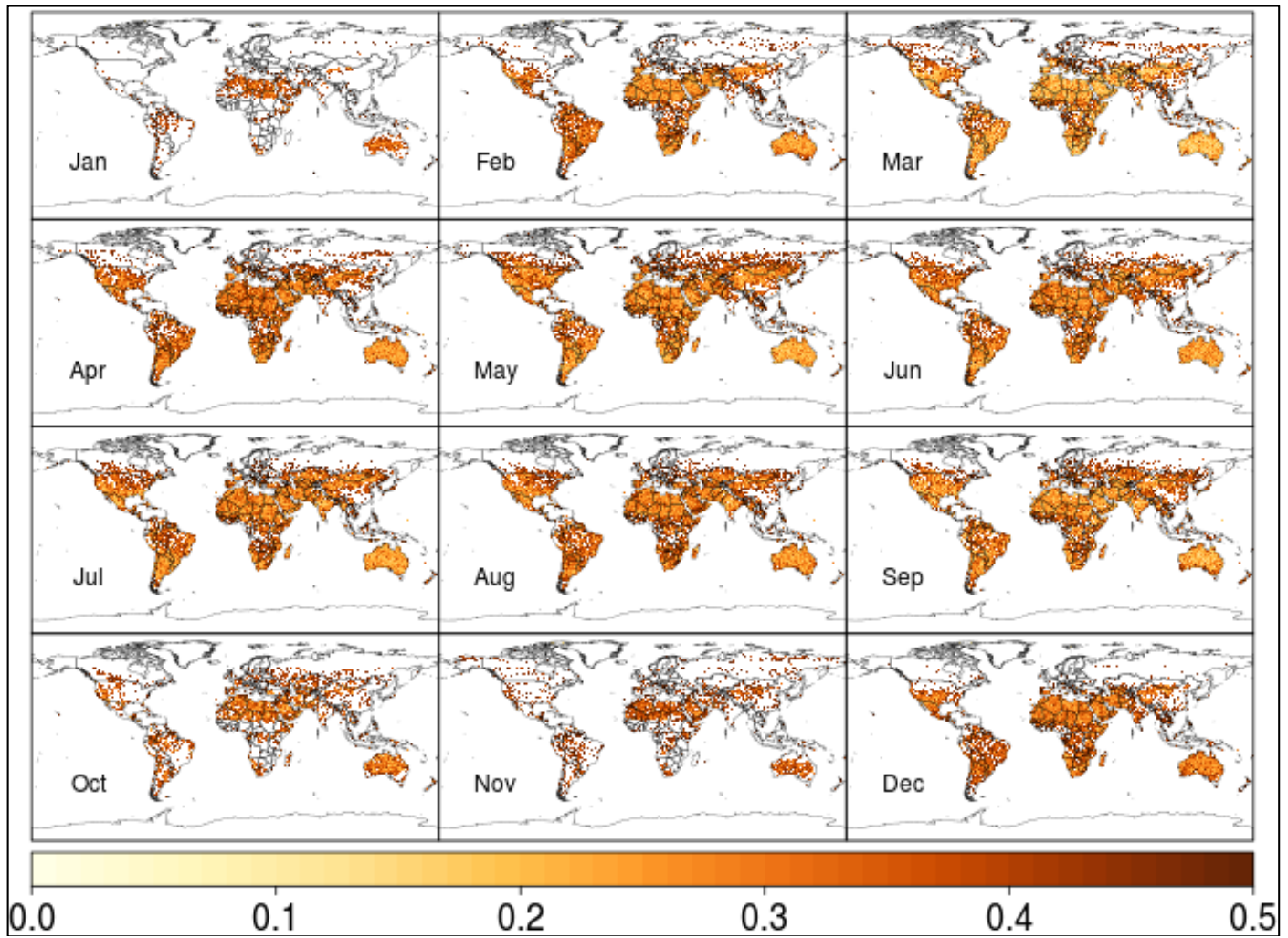


**Fig 2:** The Pixel-level means of A (IDW-EVI2), B (SG-EVI2) and C (INLA-EVI2)



**Fig 3:** The Pixel-level standard deviations of the VIP-EVI2, A = EVI2-IDW, B = SG-EVI2 and C = INLA-EVI2.

The highest level of correlation was found at mid-latitudes throughout much of the year, while average correlations were found during green-up and summer months at northernmost latitudes (March-April-May and June-July-August) and the poorest correlations were observed during brown-down (fall) months at northernmost latitudes (October-January).



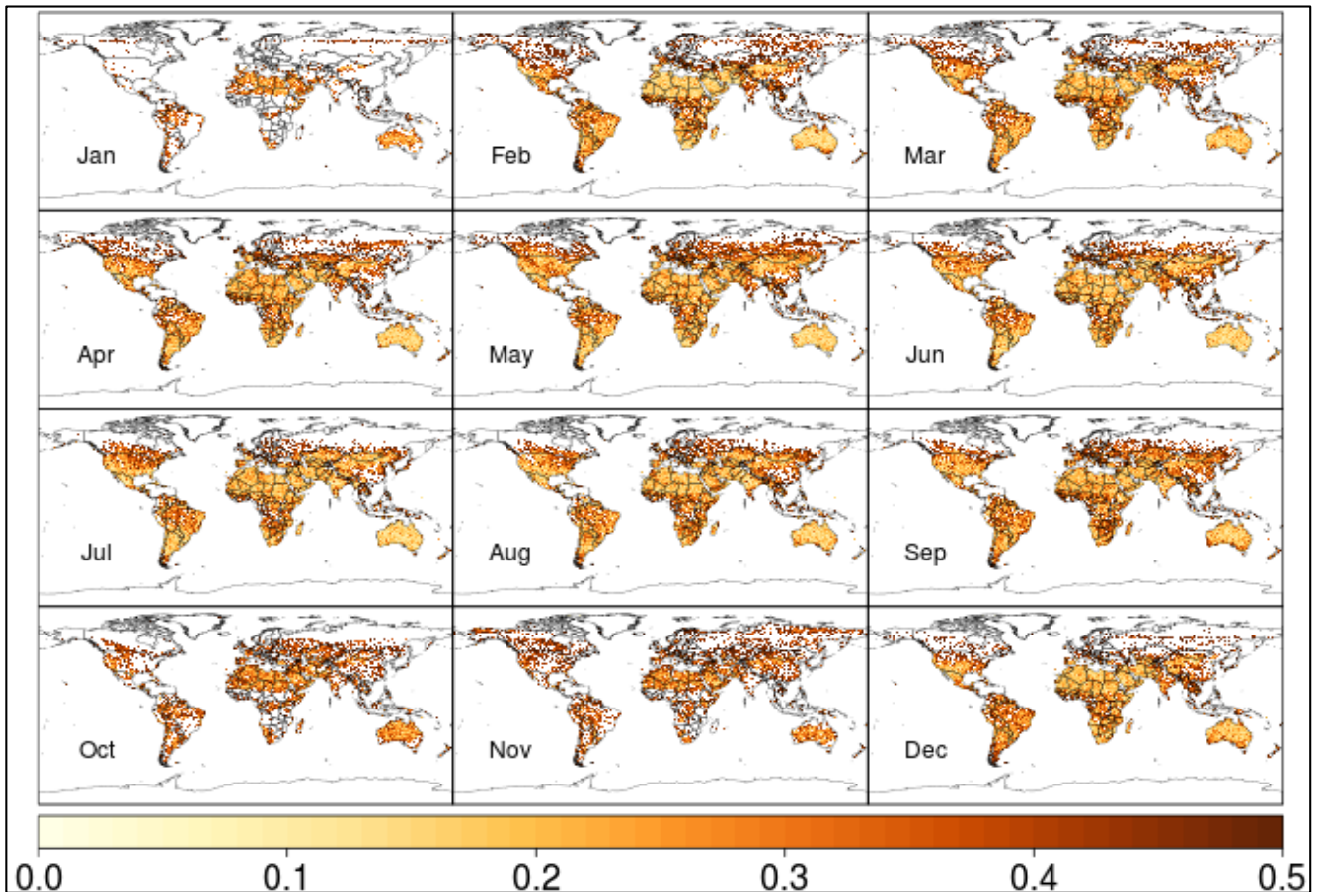
**Fig 4:** The goodness-of-fit test masked for  $D > 0.5$  based on a two sample K-S test on a per-pixel for the IDW-EVI2 versus SG-EVI2.

Overall, INLA provided better model fit with over 70% of pixels reporting  $D < 0.2$  while the SG method produced  $D$  between 0.1 and 0.4 and a majority ( $D > 0.5$ ) masked to express varied distribution.  $D > 0.5$  (varied distribution) was most prevalent during boreal and austral summer months (October-February) for both the SG-EVI2 and the INLA-EVI2 approach even though the degree of variability appears highest with the SG-EVI2.

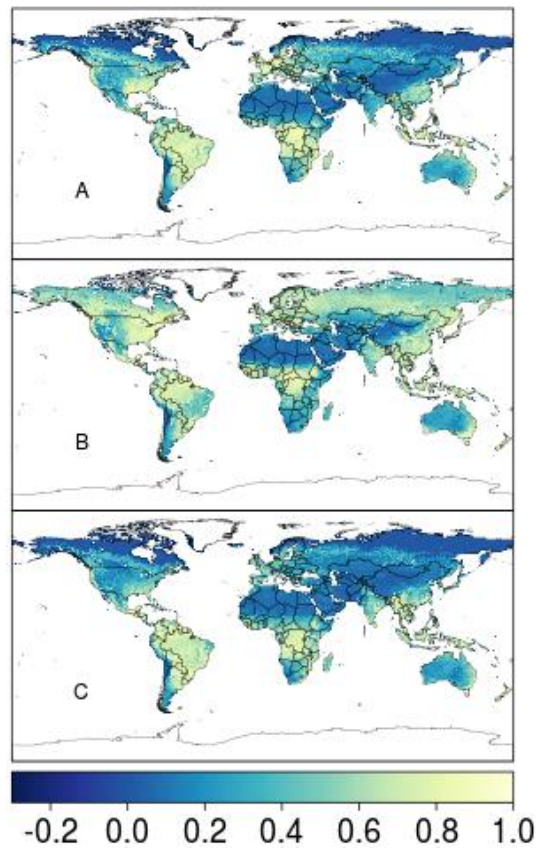
The T-S regression model masked for significance using the M-K test was used to study the annual trends in VIP-EVI2 with each technique (Figure 6). Positive values indicate a greening trend, while negative values indicate a browning trend. Overall, greening trends were more prevalent at mid-latitude and least significant at extreme high latitudes. INLA-EVI2 and IDW-EVI2 were closer in terms of trend magnitude and direction than SG-EVI2 and IDW-EVI2. The greatest difference in the techniques appear to be in their degree of sensitivity to weak browning/greening trends in which SG appear to be least sensitive as expressed by large number of insignificant trends at extreme high latitudes.

### 3.2 Local inter-comparison

Of the original 1,459 Landsat-LAI pairs, only 319 samples remained for comparison due to overlap between AVHRR and Landsat pixels in space and time. Initially, LAI and downscaled IDW-EVI2 statistics were computed for the pooled (all crops) data, but this led to poor and insignificant relationships, so statistics were computed instead on a per-crop basis. On a per-crop basis, sample sizes were small, so additional data points for some crops were omitted from the analysis, including: barely, canola, cotton, and rice. In order to increase the sample sizes of the remaining crop, two aggregations were made: alfalfa was included with pasture and garlic, onion, potato, and sugar beet were combined to form a "roots and tubers" category. After these processing steps, 225 samples remained for analysis. Based on the summary statistics used for the global comparison, INLA-EVI2 again showed the strongest correlation with LAI compared to SG-EVI2 and IDW-EVI2 (Table 1). For maize, pasture, roots and tubers, soybean, and wheat, INLA explained an additional 2%, <1%, 3%, 1%, and 1% of LAI variance than SG-EVI2, respectively. IDW showed minor improvements over SG-EVI2 for Maize ( $\Delta R^2=0.01$ ) and Roots and Tubers ( $\Delta R^2=0.02$ ), and under-performed INLA with the exception of wheat ( $\Delta R^2=0.02$ ).



**Fig 5:** The goodness-of-fit test masked for  $D > 0.5$  based on a two sample K-S test on a per-pixel for the IDW-EVI2 versus INLA-EVI2.



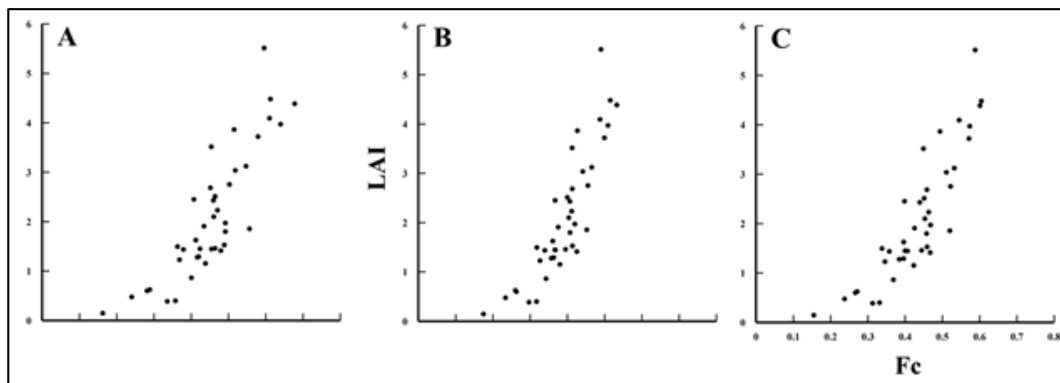
**Fig 6:** Annual Thiel-Sen regression slope: A (IDW-EVI2); B (SG-EVI2); and C (INLA-EVI2) between 1982 and 2011 masked for significance ( $\alpha = 0.05$ ).

As illustrated in Figure 7 for soybean, INLA consistently reproduced strong linear relationships over a similar range as IDW-EVI2 compared to SG-EVI2. SG-EVI2 had a lower dynamic range, which lead to a steeper slope and small, but lower correlations with *in situ* LAI.



**Table 2:** Summary statistics (N = Sample size,  $R^2$  = coefficient of determination, and RMSE = root mean squared error) from *in situ* Leaf Area Index versus linearly fitted downscaled IDW-EVI2, SG-EVI2, and INLA-EVI2. The highest performing approach is bolded.

Crop	N	Smoother	$R^2$	RMSE
Maize	108	IDW	0.63	0.934
		SG	0.62	0.944
		INLA	0.64	0.926
Pasture	35	IDW	0.50	0.558
		SG	0.51	0.554
		INLA	0.51	0.552
Roots and Tubers	19	IDW	0.91	0.545
		SG	0.89	0.604
		INLA	0.92	0.539
Soybean	40	IDW	0.72	0.697
		SG	0.74	0.669
		INLA	0.75	0.651
Wheat	23	IDW	0.52	1.091
		SG	0.49	1.128
		INLA	0.50	1.117

**Fig 7:** Scatter diagrams of *in situ* Leaf Area Index (LAI) for soybean versus downscaled IDW-EVI2 fraction ( $F_c$ ) for each smoothing technique: A (IDW-EVI2); B (SG-EVI2); and C (INLA-EVI2). LAI is expressed in units of  $m^2 m^{-2}$ .

#### 4. Discussion

This study examined the performance of a new technique to gap-fill and smooth Earth observation based long-term vegetation records with a Bayesian time series regression model using INLA. The temporal autocorrelations (trend, seasonal, and cyclical components) are adjusted for in an additive way which is unique to gap-filling and smoothing global Earth Observation based vegetation index records. The model was tested on the VIP-EVI2 version 3 record. The technique is designed to be efficient and able to handle time series where the underlying processes are not well known and contain several discontinuities and outliers. The performance of the Bayesian time series model with INLA (INLA-EVI2) was compared to the smoothed and gap-filled data provided by the VIP-EVI2 distributor (IDW-EVI2) and another commonly used technique by the remote sensing community (SG-EVI2). For both scales of analysis, INLA-EVI2 produced the highest correlations to IDW-EVI2 and *in situ* data.

Overall, the Bayesian time series model with INLA underestimated IDW-EVI2, as might be expected with posterior and/or predictive means of the data from both good quality and poor quality observations (missing values). However, the study considered predictive modes given their potential unbiasedness when reconstructing vegetation index data whose values tend to be lower than the true values due to cloud contamination. Unlike INLA-EVI2, SG-EVI2 tended to overestimate IDW-EVI2, which can be attributed to the automated optimization routine employed by the approach. According to Kandasamy *et al.* (2012) <sup>[19]</sup>, the upper envelop approach in the SG-EVI2 tends to overestimate vegetation index values in regions that are least affected by atmospheric cloud cover, because it preserves the upper envelope (VIP-EVI2 maxima) of the time series. Since the optimization routine produced a lower dynamic range than INLA-EVI2, minima were higher as well. In the global comparison, over 70% of the INLA-EVI2 and IDW-EVI2 goodness-of-fit test resulted in  $D < 0.5$ . While the distribution of D statistic for SG-EVI2 and IDW-EVI2 was comparable, it tended to result in higher D's suggesting different statistical distributions. The level of disagreement based on the D statistic ( $D > 0.5$ ) was highest at northern extremes during winter and least during summer months at mid-latitudes. The distribution of significant D statistic (disagreements not attributable to chance) was considerably higher in SG-EVI2/IDW-EVI2 than INLA-EVI2/IDW-EVI2. Coincidentally, the level of agreement between predicted values and corresponding IDW-EVI2 appear to decrease as the percentage gaps in VIP-EVI2 increases. The finding is consistent with Rusticus and Lovato (2014) <sup>[33]</sup> who noted that the overall power of equivalence testing is strongly influenced by sample size. The Lowest agreement was noted during winter months when the percentage of good quality data was lowest and highest during summer months when good quality data was most abundant. The impact of percentage gaps on reconstruction was more severe and heterogeneous for SG-EVI2 than INLA-EVI2, which is confirmed by Kandasamy *et al.* (2013) <sup>[19]</sup>, who observed that among the techniques they compared, SG was more sensitive to the number and length of data gaps. Seasonality was less apparent in the southern hemisphere, which was likely due to the greater proportion of water (i.e. less continentally).

A non-parametric regression model based on the T-S slope masked for significance with a seasonal Mann-Kendall trend test was used to determine annual trends for each smoothing technique. The technique was employed to circumvent the independence

assumption and potential effects of outliers, which in the presence of serial autocorrelation, tends to pose a challenge to matrix algebra. Despite some regional differences in areas at high latitudes with many data gaps, there was a high level of agreement between annual trends produced from the smoothing techniques compared to IDW-EVI2 in this study and those produced from GIMMS and MODIS NDVI records in Fensholt *et al.* (2012)<sup>[9]</sup> and in Marshall *et al.* (2016)<sup>[20]</sup>. Each technique produced positive (green-up) trends over much of the globe. Generally, the greening has been attributed in previous studies to an increase in and extension of primary plant production (when light and moisture are not limiting factors) in response to global warming. Despite T-S slope estimates associated with the optimization routine being consistently higher (depicting a high rate of greening), INLA-EVI2 showed more brown-down/green-up trends at extreme latitudes, while SG-EVI2 showed more negative (browning) trends which are slightly higher in absolute terms. Marshall *et al.* (2016)<sup>[20]</sup> confirms the INLA-EVI2 results, revealing that trends can be heterogeneous in both space and time, due to droughts, fires, pests, land cover change, and climate-land feedbacks. In addition, discontinuities in the data record, transitions in annual trends, and differences in the trend analysis approach used, can make comparisons particularly difficult at these latitudes.

Unique to this study, the smoothing techniques developed from coarse resolution Earth observation data were compared using ground-based data. Although the final *in situ* LAI data used for the comparison had a small sample size and poor geographical distribution, they corroborate the global comparison.  $F_C$  computed from downscaled INLA-EVI2 led to better association with *in situ* LAI and  $F_C$  computed from downscaled unsmoothed data than  $F_C$  computed from the downscaled optimization routine or IDW-EVI2. Bayesian time series model with INLA-EVI2 appear to be less impacted by outliers and discontinuities than the SG-EVI2, leading to a spread consistent with IDW-EVI2. With the exception wheat, INLA-EVI2 performed as well if not better than the IDW-EVI2 and SG-EVI2 with LAI, i.e. higher  $R^2$  and lower RMSE. The high correlations for roots and tubers should be regarded with caution, because the small sample size and leveraging in the residual plot (not shown) is most likely inflating the strength of the relationship. The moderate to high correlations across the crops and techniques suggests that the downscaling method proposed by Hwang *et al.* (2011)<sup>[15]</sup> and the *in situ* database created by Kang *et al.* (2016)<sup>[20, 23]</sup> can be used in conjunction with other *in situ* databases, such as the woody plant database (Iio *et al.*, 2014)<sup>[16]</sup> in the future, to evaluate the updated VIP version 4 and other records (e.g. MODIS) with various gap-filling and smoothing approaches.

The primary disadvantage of full Bayesian techniques especially using MCMC and a reason they have not been widely used for smoothing in the remote sensing community, is that they are computationally more intensive. The findings reveal desirable benefits that come with full Bayesian smoothing (e.g. being less severely impacted by the large fraction and length of missing data) compared to the commonly used curve-fitting techniques. INLA approach is limited to latent Gaussian models (Rue *et al.*, 2009)<sup>[32]</sup>, unlike MCMC methods which are very general, flexible and can effectively be applied to nearly any model. However, in some cases, it is difficult to assess MCMC convergence to the target posterior distribution, resulting in inference from a false positive posterior distribution. It is our view that the main benefit of the new gap-filling and smoothing procedure is computational, a short cut to full Bayesian inference with simulation based routines which is near impossible for complex models despite having desirable properties.

## 5. Conclusion

The research presented in this study highlights the important role a full Bayesian time series model with INLA can play in reconstructing Long-term Earth observation based vegetation index records where the underlying processes are not well understood and contain several outliers and discontinuities. Specifically, this study showed, using AVHRR-MODIS fused records and *in situ* data that a full Bayesian time series smoothing procedure is less severely impacted by missing data than the commonly used SG procedure.

Performance assessment with K-S test suggests more comparable findings between SG-EVI2 and INLA-EVI2 during summer but more different estimates during winter. This is reflected by the high proportion of pixels with high distance statistic ( $D > 0.5$ ) which indicates high level of disagreement with IDW-EVI2 (reference record). In particular, such pixels were more erratic during winter for SG-EVI2 but were fewer and fairly distributed across the months of the year for INLA-EVI2. The good performance of INLA-EVI2 was also demonstrated by its ability to detect long-term annual trends. While annual trend analysis with INLA-EVI2 and IDW-EVI2 were comparable across the globe, trends with SG-EVI2 were different in regions negatively impacted by atmospheric contaminations. Such Poor and largely non-significant long-term annual trends for SG-EVI2 were noted in extreme high latitude (e.g. Canada, Russia), Brazil, parts of Australia and other regions that are either water or largely Forest ecosystem (e.g. Congo). The good performance of INLA-EVI2 was further validated by its ability to correctly estimate ground observed LAI data taken for major crop areas. Of the five major crop areas (maize, pasture, roots and tubers, soybean, and wheat) considered, INLA-EVI2 predicted the *in situ* LAI data more precisely than the alternative records (SG-EVI2 and IDW-EVI2) for Maize, pasture, roots and tubers, and soybean, producing highest  $R^2$  and lowest RMSE.

Therefore, we suggest considering Bayesian time series model under INLA to gap-fill and smooth long-term Earth observation based vegetation index. The procedure is highly flexible for a large class of statistical models (<http://www.r-inla.org/>). Alternative latent models available under INLA can also be tested for the time series components to assess performance difference in terms of improved accuracy. However, if computational resources are heavily restricted, we recommend IDW and linear interpolation due to its fairly good performance. However, SG procedure could be adopted in regions less impacted by atmospheric contamination given its relatively high accuracy, simplicity and computational efficiency.

## 6. Acknowledgements

This work was supported primarily through donor contributions to the Consortium of International Research (CGIAR) Centers Research Program (CRP) on Climate Change, Agriculture and Food Security and Policies (CCAFS); and Policies, Institutions and Markets (PIM). Additional resources to download and process the VIP-EVI2 dataset were drawn from the CRP program on Forest, Trees and Agroforestry. The LAI database was developed with support from the National Aeronautics & Space

Administration Earth and Space Science Fellowship, National Science Foundation Water Sustainability & Climate Program, North Temperate Lakes long-term Ecological Research Program and University of Wisconsin-Madison Anna grant Birge Award.

## 7. References

1. Asrar G, Myneni RB, Choudhury BJ. Spatial heterogeneity in vegetation canopies and remote sensing of absorbed photosynthetically active radiation: A modeling study. *Remote Sens. Environ.* 1992; 41:85-103. doi:10.1016/0034-4257(92)90070-Z
2. Barreto-Munoz A. Multi-Sensor Vegetation Index and Land Surface Phenology Earth Science Data Records in Support of Global Change Studies: Data Quality Challenges and Data Explorer System, 2013.
3. Batista GT, Shimabukuro YE, Lawrence WT. The long-term monitoring of vegetation cover in the Amazonian region of northern Brazil using NOAA-AVHRR data. *Int. J. Remote Sens.* 1997; 18:3195-3210. doi:10.1080/014311697217044
4. Brown ME, Pinzón JE, Didan K, Morisette JT, Tucker CJ. Evaluation of the Consistency of Long-Term NDVI Time Series Derived From AVHRR, and Landsat ETM + Sensors. 2006; 44:1787-1793.
5. Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P. A knowledge-based approach to the statistical mapping of climate. *Clim. Res.* 2002; 22:99-113. doi:10.3354/cr022099
6. Didan K. Multi-satellite Earth science data record for studying global vegetation trends and changes. *Int. Geosci. Remote Sens. Symp.* 2014.
7. Fang X, Zhu Q, Chen H, Ma Z, Wang W, Song X, *et al.* Analysis of vegetation dynamics and climatic variability impacts on greenness across Canada using remotely sensed data from 2000 to 2009. *J. Appl. Remote Sens.* 2014; 8:836-66. doi:10.1117/1.jrs.8.083666
8. Fensholt R, Nielsen TT, Stisen S. Evaluation of AVHRR PAL and GIMMS 10-day composite NDVI time series products using SPOT-4 vegetation data for the African continent. *Int. J. Remote Sens.* 2006; 27:2719-2733. doi:10.1080/01431160600567761
9. Fensholt R, Proud SR. Evaluation of Earth Observation based global long term vegetation trends - Comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* 2012; 119:131-147. doi:10.1016/j.rse.2011.12.015
10. Fensholt R, Sandholt I, Rasmussen MS. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sens. Environ.* 2004; 91:490-507.
11. Friedl MA, Davis FW, Michaelsen J, Moritz MA. Scaling and uncertainty in the relationship between the NDVI and land surface biophysical variables: An analysis using a scene simulation model and data from FIFE. *Remote Sens. Environ.* 1995; 54:233-246. doi:10.1016/0034-4257(95)00156-5
12. Gitelson AA, Peng Y, Huemmrich KF. Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250m resolution data. *Remote Sens. Environ.* 2014; 147:108-120. doi:10.1016/j.rse.2014. 02.014
13. Gutman G, Ignatov A. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int. J. Remote Sens.* 1998; 19:1533-1543. doi:10.1080/014311698215333
14. Holben BN. Characteristics of maximum-value composite images from temporal AVHRR data, 1986, 37-41. doi:10.1080/01431168608948945
15. Hwang T, Song C, Bolstad PV, Band LE. Downscaling real-time vegetation dynamics by fusing multi-temporal MODIS and Landsat NDVI in topographically complex terrain. *Remote Sens. Environ.* 2011; 115:2499-2512. doi:10.1016/j.rse.2011.05.010
16. Iio A, Hikosaka K, Anten NPR, Nakagawa Y, Ito A. Global dependence of field-observed leaf area index in woody species on climate: a systematic review. *Glob. Ecol. Biogeogr.* 2014; 23:274-285.
17. Jiang Z, Huete A, Didan K, Miura T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 2008; 112:3833-3845. doi:10.1016/j.rse.2008.06.006
18. Justice CO, Eck TF, TANRÉ D, Holben BN. The effect of water vapour on the normalized difference vegetation index derived for the Sahelian region from NOAA AVHRR data. *Int. J. Remote Sens.* 1991; 12:1165-1187.
19. Kandasamy S, Neveux P, Verger A, Buis S, Weiss M, Baret F. Improving the Consistency and Continuity of MODIS 8 Day Leaf Area Index Products. *Int. J. Electron. Telecommun.* 2012; 58:141-146. doi:10.2478/v10177-012-0020-8
20. Kang Y, Özdoğan M, Zipper S, Román M, Walker J, Hong S, Marshall M *et al.* How Universal Is the Relationship between Remotely Sensed Vegetation Indices and Crop Leaf Area Index? A Global Assessment. *Remote Sens.* 2016; 8:597. doi:10.3390/rs8070597
21. Katzfuss M, Cressie N. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *J. Time Ser. Anal.* 2011; 32:430-446.
22. Lu L, Kuenzer C, Wang C, Guo H, Li Q. Evaluation of three MODIS-derived vegetation index time series for dryland vegetation dynamics monitoring. *Remote Sens.* 2015; 7:7597-7614. doi:10.3390/rs70607597
23. Marshall M, Okuto E, Kang Y, Opiyo E, Ahmed M. Global assessment of Vegetation Index and Phenology Lab (VIP) and Global Inventory Modeling and Mapping Studies (GIMMS) version 3 products. *Biogeosciences.* 2016; 13:625-639.
24. Mu Q, Heinsch FA, Zhao M, Running SW. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* 2007; 111:519-536. doi:10.1016/j.rse.2007 .04.015
25. Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y *et al.* Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* 2002; 83:214-231. doi:10.1016/S0034-4257(02)00074-3
26. Olsson L, Eklundh L, Ardö J. A recent greening of the Sahel - Trends, patterns and potential causes. *J. Arid Environ.* 2005; 63:556-566. doi:10.1016/j.jaridenv.2005 .03.008

27. Opiyo EO, Ngigi T, Nduati E, Mang M. A Prototype online system to process and visualize phenology parameters for Kenya : Case study of Bunguma county, 2013.
28. Peñuelas J, Filella L. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 1998; 3:151-156. doi:10.1016/S1360-1385(98)01213-8
29. Pinzon J, Tucker C. A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series. *Remote Sens.* 2014; 6:6929-6960. doi:10.3390/rs6086929
30. Potitsep S, Nasahara NK, Muraoka H, Nagai S, Suzuki R, Science E. What Is the Actual Relationship Between Lai and Vi in a Deciduous Broadleaf Forest? *Remote Sens. Spat. Inf. Sci.* 2010; 38:609-614.
31. Rocha AV, Shaver GR. Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes. *Agric. For. Meteorol.* 2009; 149:1560-1563. doi:10.1016/j.agrformet.2009.03.016
32. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* 2009; 71:319-392. doi:10.1111/j.1467-9868.2008.00700.x
33. Rusticus SA, Lovato CY. Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests : A Simulation Study. *Pract. Assessment, Res. Eval.* 2014; 19.
34. Sellers PJ. Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* 1985; 6:1335-1372. doi:10.1080/01431168508948283
35. Tian F, Fensholt R, Verbesselt J, Grogan K, Horion S, Wang Y. Evaluating temporal consistency of long-term global NDVI datasets for trend analysis. *Remote Sens. Environ.* 2015; 2. doi:10.1016/j.rse.2015.03.031
36. Tucker C, Pinzon J, Brown M, Slayback D, Pak E, Mahoney R, *et al.* An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.* 2005; 26:4485-4498. doi:10.1080/01431160500168686
37. Wang Q, Adiku S, Tenhunen J, Granier A. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sens. Environ.* 2005; 94:244-255. doi:10.1016/j.rse.2004.10.006
38. Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. ISPRS Journal of Photogrammetry and Remote Sensing An effective approach for gap-filling continental scale remotely sensed. *ISPRS J. Photogramm. Remote Sens.* 2014; 98:106-118. doi:10.1016/j.isprsjprs.2014.10.001
39. Weiss JL, Gutzler DS, Coonrod JEA, Dahm CN. Long-term vegetation monitoring with NDVI in a diverse semi-arid setting, central New Mexico, USA. *J. Arid Environ.* 2004; 58:249-272. doi:10.1016/j.jaridenv.2003.07.001
40. Xiao X, Braswell B, Zhang Q, Boles S, Frolking S, Moore B. Sensitivity of vegetation indices to atmospheric aerosols: Continental-scale observations in Northern Asia. *Remote Sens. Environ.* 2003; 84:385-392. doi:10.1016/S0034-4257(02)00129-3
41. Yu H, Xu J, Okuto E, Luedeling E. Seasonal response of grasslands to climate change on the Tibetan Plateau. *PLoS One.* 2012; 7:e49230. doi:10.1371/journal.pone.0049230
42. Zhou J, Jia L, Menenti M. Remote Sensing of Environment Reconstruction of global MODIS NDVI time series : Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sens. Environ.* 2015. doi:10.1016/j.rse.2015.03.018