Vegetation anomalies caused by antecedent precipitation in most of the world

Environmental Research Letters

LETTER

Vegetation anomalies caused by antecedent precipitation in most of the world

C Papagiannopoulou[1], D G Miralles[2,4], W A Dorigo[3], N E C Verhoest[2], M Depoorter[2] and W Waegeman[1]

[1] Department of Mathematical Modeling, Statistics and Bioinformatics, Ghent University, B-9000 Ghent, Belgium
[2] Laboratory of Hydrology and Water Management, Ghent University, B-9000 Ghent, Belgium
[3] Department of Geodesy and Geo-Information, Vienna University of Technology, Vienna 1040, Austria
[4] Author to whom any correspondence should be addressed.

E-mail: diego.miralles@ugent.be

Keywords: global vegetation, Granger causality, hydro-climatic extremes, water stress, ecosystem resilience

OPEN ACCESS

RECEIVED
29 June 2016

REVISED
24 April 2017

ACCEPTED FOR PUBLICATION
5 May 2017

PUBLISHED
10 July 2017

Abstract

Quantifying environmental controls on vegetation is critical to predict the net effect of climate change on global ecosystems and the subsequent feedback on climate. Following a non-linear Granger causality framework based on a random forest predictive model, we exploit the current wealth of multi-decadal satellite data records to uncover the main drivers of monthly vegetation variability at the global scale. Results indicate that water availability is the most dominant factor driving vegetation globally: about 61% of the vegetated surface was primarily water-limited during 1981–2010. This included semiarid climates but also transitional ecoregions. Intra-annually, temperature controls Northern Hemisphere deciduous forests during the growing season, while antecedent precipitation largely dominates vegetation dynamics during the senescence period. The uncovered dependency of global vegetation on water availability is substantially larger than previously reported. This is owed to the ability of the framework to (1) disentangle the co-linearities between radiation/temperature and precipitation, and (2) quantify non-linear impacts of climate on vegetation. Our results reveal a prolonged effect of precipitation anomalies in dry regions: due to the long memory of soil moisture and the cumulative, non-linear, response of vegetation, water-limited regions show sensitivity to the values of precipitation occurring three months earlier. Meanwhile, the impacts of temperature and radiation anomalies are more immediate and dissipate shortly, pointing to a higher resilience of vegetation to these anomalies. Despite being infrequent by definition, hydro-climatic extremes are responsible for up to 10% of the vegetation variability during the 1981–2010 period in certain areas, particularly in water-limited ecosystems. Our approach is a first step towards a quantitative comparison of the resistance and resilience signature of different ecosystems, and can be used to benchmark Earth system models in their representations of past vegetation sensitivity to changes in climate.

1. Introduction

Vegetation is a key player in the climate system, constraining atmospheric conditions through a series of positive and negative feedbacks. Plants regulate water, energy and carbon cycles, through their transfer of vapour from land to atmosphere (i.e. transpiration, interception loss), effects on the surface radiation budget (e.g. albedo, surface temperature, emission of volatile organic compounds), exchange of carbon dioxide with the atmosphere (i.e. photosynthesis, respiration), and influence on wind circulation (Bonan 2008, Mcpherson 2007, Teuling *et al* 2017). Vegetation holds around 42% (~28 Pg C) of the terrestrial carbon storage and assimilates about 20% of the annual anthropogenic emissions of carbon dioxide (Pan *et al* 2011, Le Quere *et al* 2016). This fundamental role highlights the importance of understanding the regional drivers of ecological sensitivity and the response of vegetation to climatic changes at the global scale.

Vegetation dynamics are generally driven by climate, in particular by precipitation, incoming radiation, air temperature and atmospheric humidity (Nemani *et al* 2003). In addition, nutrient availability (e.g. atmospheric CO_2 concentrations, soil chemicals) and short-term natural and anthropogenic disturbances (e.g. fires, volcanic eruptions, logging, insect epidemics) can be crucial at various spatiotemporal scales (Fisher *et al* 2012, Reichstein *et al* 2013, Le Quere *et al* 2016, Zhu *et al* 2016). Consequently, humans impact vegetation dynamics directly through land-use change or agricultural management, and indirectly through air pollution, induced changes in climate and spread of pest outbreaks (Baccini *et al* 2012, Reichstein *et al* 2013). In natural conditions, long-term climatological controls on vegetation dominate: this is reflected in the general distribution of continental biomes, largely based on the annual cycle of solar irradiance, mean temperature, and the intensity of dry and wet seasons (Kottek *et al* 2006). However, at shorter temporal scales, the interactions between vegetation and climate become complex and species-dependent (Zimmermann *et al* 2009). Some vegetation types react preferentially to specific climatic changes, with different levels of intensity, resilience and lagged response (Wu *et al* 2015, De Keersmaecker *et al* 2015, Seddon *et al* 2016). In addition, extreme climatic events—such as droughts, heatwaves, or heavy winds and storms—may cause long-lasting impacts and even bring ecosystems to a tipping point for collapse (Anderegg *et al* 2015, Ciais *et al* 2005, Reichstein *et al* 2013, Verbesselt *et al* 2016). Ultimately, the resistance and resilience of ecosystems to these anomalies depend on both vegetation characteristics and the duration and severity of climatic events (Anderegg *et al* 2015, Cole *et al* 2014).

A first and necessary step to understand how vegetation will respond to future climatic changes is to quantify the sensitivity of global ecosystems to past-time climate variability. Conveniently, satellites routinely collect a wealth of information about the dynamics of our biosphere, hydrosphere and atmosphere: current multi-satellite composite records of environmental and climatic variables enable the study of global vegetation–climate interactions over multi-decadal time scales. Recent studies using long-term satellite records have indicated an overall greening trend (Zhu *et al* 2016) and a long-term increase in above-ground biomass (Liu *et al* 2015)—particularly at high latitudes and in the tropics—that have been attributed to CO_2 fertilization, warming trends and land-use change. Dominant ecosystem drivers at inter- and intra-annual scales have also been intensively studied, both globally (Nemani *et al* 2003, Poulter *et al* 2014, De Keersmaecker *et al* 2015, Wu *et al* 2015, Gonsamo *et al* 2016, Seddon *et al* 2016) and over specific regions (Zhou *et al* 2014, Barichivich *et al* 2014, Guan *et al* 2015). A particular example of a well-studied phenomenon is the short-term response of the Amazonian forest to precipitation scarcity and radiation, which has been the subject of intense debate over the past few years (Morton *et al* 2014, Saleska *et al* 2016).

In the context of identifying the short-term (e.g. monthly) climatic controls on global vegetation dynamics, approaches based on correlations or multi-linear regressions between climate and vegetation variables have led to important steps forward in our understanding (Nemani *et al* 2003, Zhao and Running 2010, Wu *et al* 2015, De Keersmaecker *et al* 2015, Seddon *et al* 2016). However, these approaches are not designed to infer causality directly, and are commonly subjected to artifacts emerging from auto-correlation, non-linearity and cross-correlation between climatic drivers (Papagiannopoulou *et al* 2017). The exponential increase in the volumes of satellite, *in situ* and reanalysis records existing today, together with the consistent progress of computing science, allow for more sophisticated data-driven methods to yield robust insights into the global interactions between vegetation and climate. As such, machine-learning approaches are becoming increasingly valuable to investigate complex cause-effect relationships in geosciences, as well as to evaluate the skill of climate models in representing these interactions (Faghmous and Kumar 2014). Just recently, Papagiannopoulou *et al* (2017) adopted the well-known Granger (1969) causality framework—originally introduced in econometrics to quantify a measure of *pseudo*-causality in time series—and extended it to capture the non-linearity of vegetation–climate relationships. This was achieved by substituting the traditional linear autoregressive model used in Granger causality approaches with a non-linear random forest algorithm (Breiman 2001). This new framework has clear advantages over simpler approaches: (a) it can cope with the emerging wealth of Earth observations while preventing over-fitting, (b) it enables a robust estimation of deterministic relationships, and (c) it incorporates the non-linear nature of vegetation–climate interactions.

Here, we apply the framework by Papagiannopoulou *et al* (2017) to a large multi-dimensional data-cube of observation-based climatic and environmental records. Using this new approach to study the observed response of vegetation to radiation, temperature and water availability, we aim to uncover the sensitivity of global ecosystems to specific climatic conditions during the period 1981–2010.

2. Materials and methods

2.1. Database and feature construction

We have prioritized the use of data sets of observational nature, while the use of reanalysis data has been limited to just a few variables (see table 1). These data sets were selected from public archives, on

Table 1. Data sets used in the analysis. These data sets are used to construct predictive features for the non-linear Granger causality framework. The NDVI is used to derive the target variable. For more details, see section 2.1 and Papagiannopoulou *et al* (2017).

Variable	Data set	Primary data source	Spatial resolution (°lat./lon.)	Temporal resolution	Reference
Surface and near-surface air temperature	CRU-HR	*in situ*	0.5	monthly	Harris *et al* (2014)
	UDEL	*in situ*	0.5	monthly	Willmott and Matuura (2001)
	ISCCP	satellite	1	daily	Rossow and Dueñas (2004)
	ERA-Interim	reanalysis	0.75	3-hourly	Dee *et al* (2011)
	GISS	*in situ*	2	monthly	Hansen *et al* (2010)
	MLOST	*in situ*	5	monthly	Smith *et al* (2008)
	CFSR-Land	satellite	0.5	daily	Coccia *et al* (2015)
Precipitation	CRU-HR	*in situ*	0.5	monthly	Harris *et al* (2014)
	UDEL	*in situ*	0.5	monthly	Willmott and Matuura (2001)
	CMAP	satellite/*in situ*	2.5	monthly	Xie and Arkin (1997)
	CPC-U	*in situ*	0.5	daily	Xie *et al* (2007)
	GPCC	*in situ*	0.5	monthly	Schneider *et al* (2008)
	GPCP	satellite/*in situ*	2.5	monthly	Adler *et al* (2003)
	ERA-Interim	reanalysis	1	daily	Dee *et al* (2011)
	MSWEP	satellite/*in situ*/reanalysis	0.25	3-hourly	Beck *et al* (2017)
Incoming short/longwave, and surface net radiation	SRB	satellite	1	3-hourly	Stackhouse *et al* (2004)
	ERA-Interim	reanalysis	0.75	3-hourly	Dee *et al* (2011)
Surface and root-zone soil moisture	GLEAM	satellite	0.25	daily	Miralles *et al* (2011)
	ESA CCI-Passive	satellite	0.25	daily	Dorigo *et al* (2017)
	ESA CCI Combined	satellite	0.25	daily	Liu *et al* (2012)
Snow water equivalent	GLOBSNOW	satellite	0.25	daily	Luojus *et al* (2013)
NDVI	GIMMS NDVI3g	satellite	0.25	monthly	Tucker *et al* (2005)

the basis of meeting a series of spatiotemporal requirements: (a) spanning multi-decadal records, (b) covering the entire vegetated continental domain, and (c) being available at an adequate spatial and temporal resolution. They span a common 1981–2010 period for the entire continental surface, and were resampled to a monthly temporal scale and 1° latitude-longitude spatial resolution using arithmetic means (for finer resolution data sets) and linear interpolation (for coarser ones). Different climatic and environmental drivers have been considered: land surface and near-surface air temperature, surface incoming (longwave and shortwave) and surface net radiation, precipitation, surface and root-zone soil moisture, and snow water equivalent. Rather than using a single data set for each of these variables, we collected the largest possible number of data sets meeting the above-mentioned requirements. This yielded a total of twenty-one different data sets (table 1).

Then, several predictive features were constructed from these 21 data sets, to be used later as predictors in our framework. These predictive features consist of monthly time series for each 1° pixel, and include: raw data time series of each data set, seasonal anomalies (after subtraction of the seasonal cycle based on the multi-year mean for each corresponding month of the year), de-trended seasonal anomalies (after subtraction of the long-term linear trend from the seasonal anomalies), lagged variables (with monthly lags up to six months into the past), cumulative variables (corresponding to the cumulative mean over the antecedent one to six months), and extreme indices (including the maximum and minimum of a variable per month, number of days per month exceeding a given threshold, values of specific percentiles, etc.). The lagged variables, cumulative variables and extreme indices were computed based on both raw data and (de-)trended seasonal anomalies. This procedure yielded a total of 3197 predictive features —see Papagiannopoulou *et al* (2017) for details.

Finally, these predictive features are used to train the non-linear Granger causality framework (see section 2.2), using as target variable the de-trended Normalized Difference Vegetation Index (NDVI) seasonal anomalies from the Global Inventory Modeling and Mapping Studies (GIMMS) 3g data set (Tucker *et al* 2005). This allows us quantify the importance of each climate feature (or group of features) for the occurrence of past vegetation (NDVI) anomalies at each pixel (see section 2.3). While other vegetation indices are being explored for their use as target variable, GIMMS NDVI is applied here for being the most widely used long-term record of vegetation greenness (Beck *et al* 2011).

2.2. Non-linear Granger causality framework

Given a particular target time series, one speaks of the existence of 'Granger causality' if the prediction of this target variable improves when information from other time series is taken into account in this prediction (Granger 1969). Here, we quantify the extent to which a variable x (i.e. a predictive feature, or a certain group of them—see section 2.3) is 'Granger-causing' a target variable y (i.e. the de-trended NDVI anomalies at each individual pixel) by computing the increase in the variance of y that is explained by the random forest model predictions when x is included in the set of predictive features used by the model (this set also includes past values of y to conform to the definition of Granger causality). The explained variance is then defined as $R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$, with RSS being the sum of squared errors of the predictions (relative to the true de-trended NDVI anomalies), and TSS being the sum of the squared differences between the true values and their long-term mean.

The model typically used to generate predictions in Granger causality analysis is a linear vector autoregressive model. As mentioned above, here we apply the extension described in Papagiannopoulou *et al* (2017) where this linear model is substituted with a more complex random forest. Generically, the random forest model is a non-linear machine-learning algorithm that emerges from a combination of multiple decision trees (Breiman 2001). As such, our Granger-causality framework allows for the identification of non-linear deterministic relationships between climate and vegetation. The implications and advantages of using this approach are extensively discussed in Papagiannopoulou *et al* (2017).

2.3. Sequential method to evaluate the impact of specific groups of features

To explore the importance of different climatic variables for the occurrence of NDVI anomalies, all 3197 predictive features have been aggregated into one of these three groups: 'temperature' (including surface and air temperature), 'radiation' (including incoming shortwave, longwave and net) and 'water' (including precipitation, surface and root-zone soil moisture, and snow water equivalent)—see table 1. Then, taking the 'water' group as example, the explained variance (R^2) of NDVI anomalies by 'water' is calculated sequentially by: (1) applying the random forest approach to predict the anomalies of NDVI based on the entire database of predictors (including 'water', 'radiation' and 'temperature' features, but also past NDVI values to conform to the definition of Granger causality); (2) applying the random forest approach to the entire database except for the 'water' group; (3) calculating the deterioration in the predictive performance (R^2) after excluding the 'water' group. Moreover, to prevent favouring groups with a larger number of predictive features, the number of selected features in every random forest is forced to be the same for all three groups.

As in most statistical techniques, the assessment of causality is ultimately limited to quantifying the level of cross-covariance between predictors and target variable, thus if critical predictors are not included, the importance of the assessed variable (or group of variables) may be inflated. However, the sequential approach explained above preserves the multivariate nature of the framework, as opposed to a hypothetical case in which the contribution of a specific group of variables (e.g. 'water') is assessed in isolation. In addition, the approach goes one step beyond previous statistical analyses of global vegetation drivers by preventing the importance of a secondary driver (e.g. temperature) to be inflated due to its correlation to the primary one (e.g. water availability). Nonetheless, the resulting R^2 is still not a measure of 'real' causality but of *pseudo*-causality, given the unfeasibility of including all possible drivers of global vegetation. Finally, we note that since the R^2 attributed to a particular variable (or group) is quantified by subtracting it from the entire database of predictors, its causal effect (computed as R^2) will be underestimated as long as the remaining variables are strongly correlated to that one being subtracted. As such, the R^2 reported here refers to the explained variance that is 'unique' to the variable (or group), i.e. the part of the variance in the NDVI anomalies that cannot be explained by any other variable in the database.

3. Results and discussion

More than half (61%) of the vegetated area appears primarily controlled by water availability (i.e. precipitation, soil moisture or snow dynamics)—see figure 1 (*a*). In addition, for 17% of the remaining vegetated area, water availability is the second most important limiting factor after temperature or radiation (figure 1 (*b*)). Temperature and radiation are the primary climatic controls for 23% and 15% of the vegetated areas, respectively. In addition, for most of these energy-driven regions the dynamics of vegetation are largely independent from climate variability (figure 1 (*b*)). That is the case for both high latitudes and tropical zones, where no climatic driver is responsible for a substantial fraction of the variability in vegetation, as suggested by the inability of the framework to explain the dynamics in NDVI anomalies (see below). Nonetheless, in boreal and temperate regions, radiation and temperature remain the two main climatic controls, respectively, and most European croplands are temperature-driven (figure 1 (*a*) and (*b*)). The relative importance of water availability in boreal regions such as Siberia or Alaska responds to the influence of snowmelt, which is nonetheless controlled by temperature and radiation patterns (Barichivich *et al* 2014). Meanwhile, central Europe is mostly temperature-driven, while China is largely controlled by radiation. These patterns are in

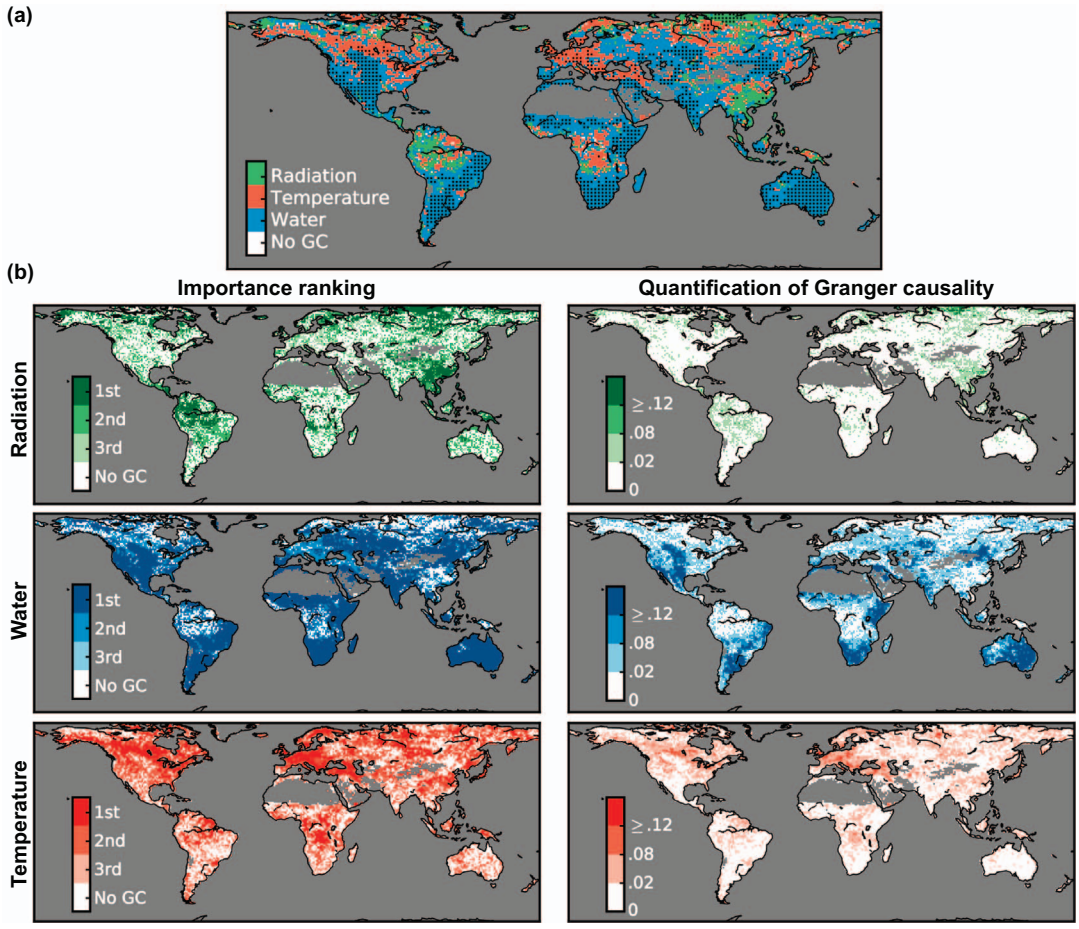

Figure 1. Primary climatic and environmental factors controlling vegetation dynamics. (*a*) Temperature, radiation and water-limited continental regions based on the non-linear Granger causality approach targeting de-trended NDVI anomalies. The net of black dots is represented at 2 degree resolution and indicates areas with $R^2 > 0.3$ for the full model including all variables. 'No GC' indicates no Granger causality ($R^2 \leq 0$). (*b*) Order of importance of each group of variables for vegetation according to the performance in terms of R^2 (left), and the corresponding R^2 (right). Grey colour indicates the regions considered as non-vegetated throughout the analysis.

general agreement with those by Nemani *et al* (2003), Wang *et al* (2011), Wu *et al* (2015) and Seddon *et al* (2016), bearing into consideration the different periods and seasons of focus, and the differences in methodology and data.

For the remaining vegetated land, the availability of water is the first control over ecosystem dynamics, and is particularly important in semiarid regions such as eastern and central Australia, the Pampas and Caatinga region in South America, the US Great Plains, and the south and Horn of Africa. Interestingly, most of these ecoregions were recently shown to influence their own availability of water through transpiration feedbacks during dry and wet years (Miralles *et al* 2016). As mentioned above, in tropical forests, none of the climatic drivers is causing a large fraction of vegetation variability. This may be explained by the subtle changes in vegetation and the ecosystem's resistance to mean climate dynamics. However, this low response of tropical rainforests may also reflect aspects such as the dependency on phenological processes driven by biotic factors (Hutyra *et al* 2007), occurrence of wild fires (van der Werf *et al* 2008), limitations imposed by the availability of soil nutrients (Fisher *et al* 2012) and tropical deforestation (Hansen *et al* 2013). In addition, it may also echo the influence of CO_2 fertilization, even though CO_2 emissions are expected to be more important for multi-decadal trends than for monthly dynamics (Liu *et al* 2015, Zhu *et al* 2016). Therefore, figure 1 only partially supports the hypothesis of a radiation constraint on tropical vegetation, as defended by Nemani *et al* (2003) or Seddon *et al* (2016): (a) the Amazonian rainforest is affected by radiation, yet the South East Asia and Congo rainforests appear primarily driven by temperature; (b) other (non-climatic) drivers seem to dominate the dynamics in these ecosystems, as discussed above. Nevertheless, known issues of NDVI saturation in densely vegetated areas (Beck *et al* 2011) may contribute to these results and should be considered.

The primary climatic controls over vegetation dynamics may shift throughout the year, both due to natural phenological cycles as well as intra-annual climate variability. In figure 2 our framework is applied to estimate the dominant factors causing vegetation variability during two distinct six-month seasons: January–June and July–December. As

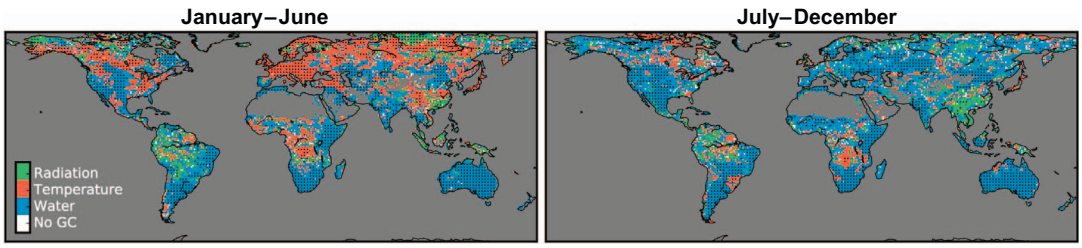

Figure 2. Factors controlling vegetation dynamics for two annual seasons. Temperature, radiation and water-limited regions during January–June (left) and July–December (right). The net of black dots is represented at 2 degree resolution and indicates areas with $R^2 > 0.3$. 'No GC' indicates no Granger causality ($R^2 \leq 0$). Grey colour indicates the regions considered as non-vegetated throughout the analysis.

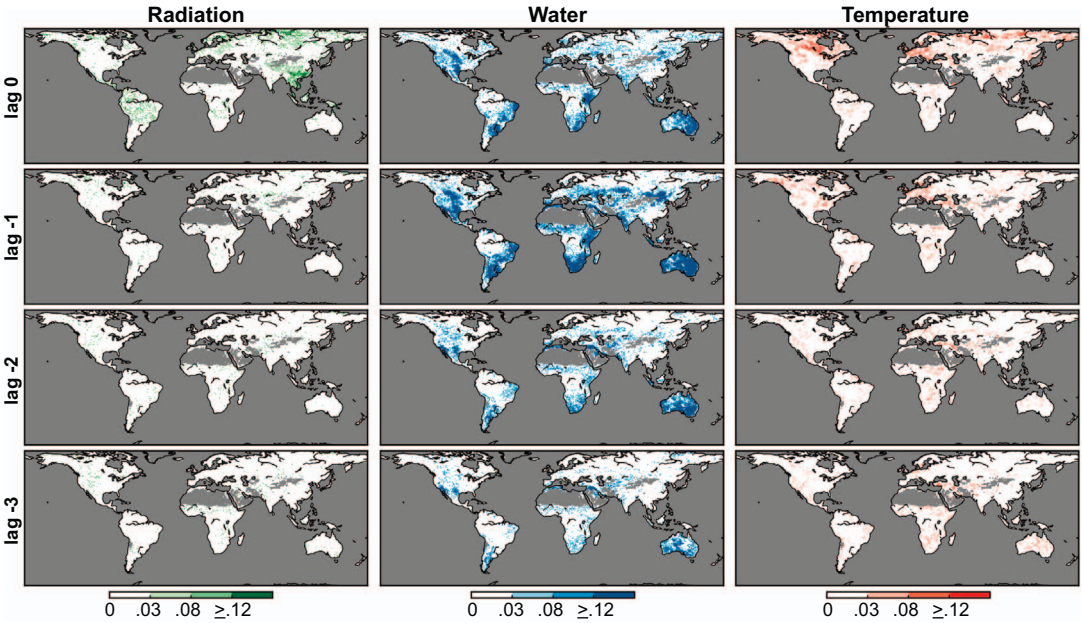

Figure 3. Temporal scale of the effects of hydro-climatic variables on vegetation. Influence (R^2) of each group of variables (radiation, temperature and water availability) on the NDVI anomalies considering different lag times (in months) separately.

expected, results are markedly different in regions of ample phenological cycles, such as Northern Hemisphere mid and high latitudes. Deciduous and mixed forests in North America, Europe and China show a strong dependency on temperature during January–June, which is consistent with the expectations of the timing and length of their growing season being dependent on temperature (see e.g. Chmielewski and Rötzer 2002, Menzel and Fabian 1999). On the other hand, precipitation occurring during summer and autumn appears more relevant as a control of Northern Hemisphere deciduous vegetation during the senescence period (Xie *et al* 2015)—see results for July–December in figure 2. Consequently, 50% of the vegetated surface appears primarily water-limited during January–June, while a larger 66% is primarily water-limited during July–December, with this seasonal dependency being mainly attributed to the phenological cycle of deciduous forests.

Since vegetation, soil and atmosphere have a memory, and because some vegetation properties take time to respond to environmental changes, it is crucial to explore the latency in this response, which is ultimately related to the resistance and resilience of the ecosystems. Lag times are already considered in figure 1 and 2, given that our non-linear approach includes predictive features with various lags and based on several past cumulative periods (see section 2.1 and Papagiannopoulou *et al* 2017). While the concept of introducing lag times in the study of these relationships is certainly not new (Davis 1984, Braswell *et al* 1997), it has become more extended in recent years (Chen *et al* 2014, Wu *et al* 2015, Seddon *et al* 2016). The aforementioned studies suggest that the time taken by vegetation to respond to climatic and environmental anomalies, as well as its resilience, depend on both climate and ecosystem characteristics. Figure 3 shows that changes in water availability lead to lagged effects and longer-term impacts on vegetation than those in radiation and temperature. Semiarid ecosystems in Australia and the Americas show sensitivity to the dynamics in water availability occurring even longer than three months earlier, which partly reflects their lower resilience to drought

stress (De Keersmaecker *et al* 2015). In addition, figure 3 confirms that vegetation greenness typically takes several weeks to react to precipitation anomalies (Adegoke and Carleton 2002, Seddon *et al* 2016): the available water during the previous month (i.e. lag = −1) has more predictive power than during the current month (lag = 0).

On the other hand, the effect of temperature and radiation is more immediate (maximum at lag = 0), and dissipates rapidly, indicating a higher resilience of vegetation to anomalies in these variables. This is supported by the results in figure 3, which show that temperature and radiation data cannot help predict vegetation greenness in the following month, not even in energy-limited regions. These results also relate to the short memory of atmosphere compared to that of soil, implying that air temperature and radiation anomalies are less likely to prevail than those of water availability in following months (Seneviratne *et al* 2006). These insights from figure 3 agree with the results by Seddon *et al* (2016), but disagree with Wu *et al* (2015). The latter showed a delayed response of vegetation greenness to radiation anomalies, based on multi-linear regressions and a partial correlation model. However, as mentioned in section 2.3, multi-linear regressions are prone to inflate the importance of temperature and radiation due to the (negative) correlations these variables hold against precipitation and soil moisture (Papagiannopoulou *et al* 2017). More complex frameworks, such as the one proposed here, allow us to disentangle and quantify the impacts of different climatic drivers independently and deterministically, which seems a necessary step to advance our understanding on climate–vegetation interactions.

Finally, we specifically target the net effect of hydro-climatic events—i.e. extremes in temperature, water availability and radiation—on global vegetation. Recent studies have highlighted the key role such events play for the structure and functioning of ecosystems, with their impacts depending on timing, magnitude, extent and type of event, and on the natural resistance and resilience of the ecosystem (Reichstein *et al* 2013, Zscheischler *et al* 2014, Sippel *et al* 2016). Because our database of predictors includes climate extreme indices calculated based on the data sets in table 1 (see section 2.1 and Papagiannopoulou *et al* 2017), we have the means to isolate the importance of hydro-climatic extremes for global ecosystems following the sequential approach described in section 2.3. Figure 4 depicts this importance in terms of R^2, i.e. the added explanatory power of these climate extremes—over the remaining predictor variables in the database—when it comes to predicting past NDVI anomalies. Hydrological extremes had an influence over the vegetation dynamics in most ecoregions on Earth during 1981–2010, being more important in areas such as the US and Australia, where severe droughts

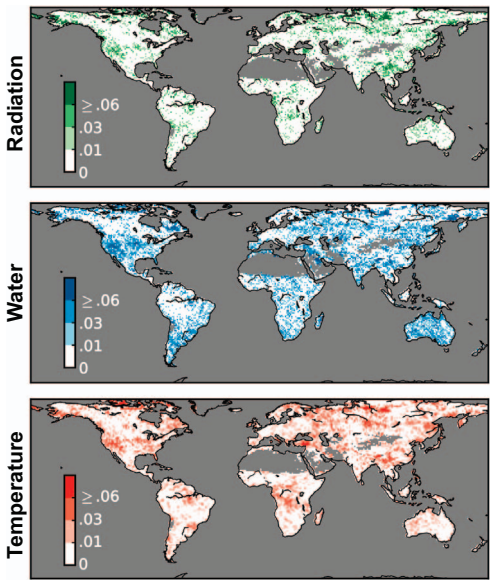

Figure 4. Effect of hydro-climatic extremes on vegetation. Influence (R^2) of radiation, water and temperature extremes on vegetation, calculated as their potential to predict the de-trended NDVI anomalies during the period 1981–2010.

occurred in recent decades. As expected, radiation and temperature extremes have an impact at higher latitudes and in the tropics; in particular, parts of boreal and tropical forests were affected by high temperature events. The apparent response of boreal forests to extremes in temperature is in line with the results by Zscheischler *et al* (2014). Despite a particular type of extreme being able to explain up to 5%–10% of past vegetation variability for some regions, the importance of these events is low compared to that of the general climate dynamics. This is simply related to the fact that extremes are by definition infrequent, thus for the multi-decadal period considered here vegetation typically responds to regular environmental conditions.

Despite the general agreement of our results with previous literature, an overall finding emerges from our analysis: water availability is not only the dominant control factor over vegetation in semiarid regions, but in most transitional ecoregions as well. On the contrary, Wu *et al* (2015) reported that most of the vegetated land is primarily controlled by temperature, then radiation and finally water (the latter accounting only for 16% of the area where results were significant), while Nemani *et al* (2003) reported 40%, 33% and 27% of the vegetated land being primarily constrained by water, temperature and radiation, respectively. Here, we estimate a contrasting 61%, 23% and 15%, which is qualitatively more comparable to the results by Seddon *et al* (2016). The latter reported water limitations in regions that were predominantly energy-limited according to Nemani *et al* (2003), such as Western Europe and the American prairies. Figure 3 supports the hypothesis that (a) our consideration of the non-linear, lagged and cumulative impacts of water availability on vegetation, (b) the treatment

of the co-linearities between temperature, radiation, and precipitation by our Granger-causality model (section 2.2), and (c) sequential approach to unravel the importance of these separate drivers (section 2.3), are behind the stronger importance of water availability for vegetation dynamics revealed in our study. Nonetheless, it is important to note that differences in the accuracy of the radiation, water and temperature observations used here could affect the resulting contributions of these drivers, and may explain part of the differences with previous studies.

4. Conclusion

We have identified the main climatic and environmental controls on global vegetation during the satellite era following a non-linear Granger causality framework, which uses random forests as core model and is driven by a large database of global observational features (Papagiannopoulou *et al* 2017). Results indicate that water availability is the primary factor driving NDVI anomalies globally, with 61% of the vegetated continental surface being water-limited, despite the relative importance of temperature in the Northern Hemisphere during the growing season. This overall water constraint appears more dominant than previously reported (Nemani *et al* 2003, Wu *et al* 2015, Gonsamo *et al* 2016). In semiarid environments, water control over vegetation is reinforced by the long memory of soil moisture, which allows precipitation to affect vegetation dynamics more than three months into the future, in contrast to the more immediate and shorter-lasting impacts of radiation and temperature. We argue that this kind of non-linear interactions have not been adequately exposed by more traditional studies based on correlations and multi-linear regression models.

Overall, our findings highlight a strong dependency of global vegetation on water availability, and show the imprint of hydro-climatic extremes on global vegetation during the satellite era. These results suggest that over a large part of the continents vegetation is prone to follow future trends in water availability. Critically, for most of the regions reported here as water-limited, the supply of precipitation is expected to decline following global warming (Fischer *et al* 2014), and a general aggravation in hydro-climatic extremes is also expected (Seneviratne *et al* 2012, Fischer *et al* 2014). In the light of these projections, further studies to characterize the resistance and resilience of global vegetation to precipitation scarcity remain imperative to adequately predict the fate of these ecosystems.

Author contributions

D G Miralles, W Waegeman and N E C Verhoest conceived the study. C Papagiannopoulou conducted the analysis. D G Miralles and C Papagiannopoulou led the writing. All co-authors contributed to the design of the experiments, interpretation of results, and editing of the manuscript.

Acknowledgments

This work is funded by the Belgian Science Policy Office (BELSPO) in the framework of the STEREO III programme, project SAT-EX (SR/00/306). D G Miralles acknowledges support from the European Research Council (ERC) under grant agreement n° 715254 (DRY–2–DRY). W A Dorigo is supported by the TU Wien Science Award 2015. We sincerely thank Stijn Decubber and Matthias Demuzere for the feedback and discussions. We thank the individual developers of the wide range of global data sets used in this study. Data and codes can be retrieved using www.sat-ex.ugent.be as gateway.

References

Adegoke J O and Carleton A M 2002 Relations between soil moisture and satellite vegetation indices in the US Corn Belt *J. Hydrometeorol.* **3** 395–405

Anderegg W R L, Schwalm C, Biondi F and Camarero J J 2015 Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models *Science* **349** 528–32

Adler R F *et al* 2003 The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present) *J. Hydrometeorol.* **4** 1147–67

Baccini A *et al* 2012 Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps *Nat. Clim. Change* **2** 182–5

Barichivich J *et al* 2014 Temperature and snow-mediated moisture controls of summer photosynthetic activity in northern terrestrial ecosystems between 1982 and 2011 *Remote Sens.* **6** 1390–431

Beck H E, McVicar T R, van Dijk A I J M, Schellekens J, de Jeu R A M and Bruijnzeel L A 2011 Global evaluation of four AVHRR–NDVI data sets: intercomparison and assessment against Landsat imagery *Remote Sens. Environ.* **115** 2547–63

Beck H E, van Dijk A I, Levizzani V, Schellekens J, Miralles D G, Martens B and de Roo A 2017 MSWEP: 3-hourly 0.25 global gridded precipitation 1979–2015 by merging gauge, satellite, and reanalysis data *Hydrol. Earth Syst. Sci.* **21** 589–615

Bonan G B 2008 Forests and climate change: forcings, feedbacks, and the climate benefits of forests *Science* **320** 1444–9

Braswell B H, Schimel D S, Linder E and Moore B 1997 The response of global terrestrial ecosystems to interannual temperature variability *Science* **248** 870–3

Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

Chen T, de Jeu R A M, Liu Y Y, van Der Werf G R and Dolman A J 2014 Using satellite based soil moisture to quantify the water driven variability in NDVI: a case study over mainland Australia *Remote Sens. Environ.* **140** 330–8

Chmielewski F-M and Rötzer T 2002 Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes *Clim. Res.* **19** 257–64

Ciais P *et al* 2005 Europe-wide reduction in primary productivity caused by the heat and drought in 2003 *Nature* **437** 529–33

Coccia G, Siemann A L, Pan M and Wood E F 2015 Creating consistent datasets by combining remotely-sensed data and land surface model estimates through Bayesian uncertainty post-processing: the case of land surface temperature from HIRS *Remote Sens. Environ.* **170** 290–305

Cole L E S, Bhagwat S A and Willis K J 2014 Recovery and resilience of tropical forests after disturbance *Nat. Commun.* **5** 1–7

Davis M B 1984 Climatic instability, time, lags, and community disequilibrium *Community Ecology* ed J M Diamond and T J Case (New York: Harper and Row) pp 269–28

Dee D P *et al* 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system *Q. J. R. Meteorol. Soc.* **137** 553–97

De Keersmaecker W, Lhermitte S, Tits L, Honnay O, Somers B and Coppin P 2015 A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover *Glob. Ecol. Biogeogr.* **24** 539–48

Dorigo *et al* 2017 ESA CCI soil moisture for improved earth system understanding: state-of-the-art and future directions *Remote Sens. Environ.* submitted

Faghmous J H and Kumar V 2014 A big data guide to understanding climate change: the case for theory-guided data science *Big Data* **2** 155–63

Fischer E M, Sedláček J, Hawkins E and Knutti R 2014 Models agree on forced response pattern of precipitation and temperature extremes *Geophys. Res. Lett.* **41** 8554–62

Fisher J B, Badgley G and Blyth E 2012 Global nutrient limitation in terrestrial vegetation *Glob. Biogeochem. Cycles* **26** GB3007

Gonsamo A, Chen J M and Lombardozzi D 2016 Global vegetation productivity response to climatic oscillations during the satellite era *Glob. Change Biol.* **22** 3414–26

Granger C W 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica* **37** 424–38

Guan K *et al* 2015 Photosynthetic seasonality of global tropical forests constrained by hydroclimate *Nat. Geosci.* **8** 284–9

Hansen J, Ruedy R, Sato M and Lo K 2010 Global surface temperature change *Rev. Geophys.* **48** RG4004

Hansen M C *et al* 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–3

Harris I P, Jones P D, Osborn T J and Lister D H 2014 Updated high-resolution grids of monthly climatic observations-the CRU TS3.10 Dataset *Int. J. Climatol.* **34** 623–42

Hutyra L R, Munger J W, Saleska S R, Gottlieb E, Daube B C, Dunn A L, Amaral D F, de Camargo P B and Wofsy S C 2007 Seasonal controls on the exchange of carbon and water in an Amazonian rain forest *J. Geophys. Res.* **112** G03008

Kottek M, Grieser J, Beck C, Rudolf B and Rubel F 2006 World map of the Koppen–Geiger climate classification updated *Meteorol. Z.* **15** 259–63

Le Quere C *et al* 2016 Global carbon budget 2016 *Earth System Science Data* **8** 605–49

Liu Y Y *et al* 2012 Trend-preserving blending of passive and active microwave soil moisture retrievals *Remote Sens. Environ.* **123** 280–97

Liu Y Y, van Dijk A I J M, de Jeu R A M, Canadell J G, McCabe M F, Evans J P and Wang G 2015 Recent reversal in loss of global terrestrial biomass *Nat. Clim. Change* **5** 470–4

Luojus K, Pulliainen J, Takala M, Lemmetyinen J, Kangwa M, Smolander T and Derksen C 2013 Global snow monitoring for climate research: algorithm theoretical basis document (ATBD)—SWE-algorithm, *Tech. Rep.* Version/Revision 1.0/02

Menzel A and Fabian P 1999 Growing season extended in Europe *Nature* **397** 659

Mcpherson R A 2007 A review of vegetation–atmosphere interactions and their influences on mesoscale phenomena *Prog. Phys. Geogr.* **31** 261–85

Miralles D G, Holmes T R, De Jeu R A, Gash J H, Meesters A G and Dolman A J 2011 Global land-surface evaporation estimated from satellite-based observations *Hydrol. Earth Syst. Sci.* **15** 453–69

Miralles D G, Nieto R, McDowell N G, Dorigo W A, Verhoest N E, Liu Y Y, Teuling A J, Dolman A J, Good S P and Gimeno L 2016 Contribution of water-limited ecoregions to their own supply of rainfall *Environ. Res. Lett.* **11** 124007

Morton D C, Nagol J, Carabajal C C, Rosette J, Palace M, Cook B D, Vermote E F, Harding D J and North P R J 2014 Amazon forests maintain consistent canopy structure and greenness during the dry season *Nature* **506** 221–4

Nemani R R *et al* 2003 Climate-driven increases in global terrestrial net primary production from 1982 to 1999 *Science* **300** 1560–3

Pan Y *et al* 2011 A large and persistent carbon sink in the world's forests *Science* **333** 988–93

Papagiannopoulou C, Miralles D G, Verhoest N E C, Dorigo W A and Waegeman W 2017 A non-linear Granger causality framework to investigate climate–vegetation dynamics *Geosci. Model Dev.* **10** 1945–60

Poulter B *et al* 2014 Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle *Nature* **509** 600–4

Reichstein M *et al* 2013 Climate extremes and the carbon cycle *Nature* **500** 287–95

Rossow W B and Dueñas E N 2004 The international satellite cloud climatology project (ISCCP) web site: an online resource for research *Bull. Am. Meteorol. Soc.* **85** 167–72

Saleska S R, Wu J, Guan K, Araujo A C, Huete A, Nobre A D and Restrepo-Coupe N 2016 Dry-season greening of Amazon forests *Nature* **531** E4–5

Schneider U *et al* 2011 GPCC full data reanalysis version 6.0 at 0.5°: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data (Global Precipitation Climatology Centre) (https://doi.org/10.5676/DWD_GPCC/FD_M_V7_050)

Seddon A W R, Macias-Fauria N, Long P R, Benz D and Willis K J 2016 Sensitivity of global terrestrial ecosystems to climate variability *Nature* **531** 229–32

Seneviratne S I *et al* 2006 Soil moisture memory in AGCM simulations: analysis of global land-atmosphere coupling experiment (GLACE) data *J. Hydrol.* **7** 1090–112

Seneviratne S I *et al* 2012 Changes in climate extremes and their impacts on the natural physical environment *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* ed C B Field *et al* (Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)) 109–230 (Cambridge: Cambridge University Press)

Sippel S, Zscheischler J and Reichstein M 2016 Ecosystem impacts of climate extremes crucially depend on the timing *Proc. Natl Acad. Sci.* **113** 5768–70

Smith T M, Reynolds R W, Peterson T C and Lawrimore J 2008 Improvements to NOAA's historical merged land-ocean surface temperature analysis 1880–2006 *J. Clim.* **21** 2283–96

Stackhouse J W, Gupta S K, Cox S J, Mikovitz C, Zhang T and Chiacchio M 2004 12-Year surface radiation budget *GEWEX News* **14** 10–2

Teuling A J *et al* 2017 Observational evidence for cloud cover enhancement over western European forests *Nat. Commun.* **8** 14065

Tucker C J, Pinzon J E, Brown M E, Slayback D A, Pak E W, Mahoney R, Vermote E F and Saleous N E 2005 An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data *Int. J. Remote Sens.* **26** 4485–98

van der Werf G R, Randerson J T, Giglio L, Gobron N and Dolman A J 2008 Climate controls on the variability of fires in the tropics and subtropics *Glob. Biogeochem. Cycles* **22** 330–8

Verbesselt J *et al* 2016 Remotely sensed resilience of tropical forests *Nat. Clim. Change* **6** 1028–31

Wang X, Piao S, Ciais P, Li J, Friedlingstein P, Koven C and Chen A 2011 Spring temperature change and its implication in the change of vegetation growth in North America from 1982 to 2006 *Proc. Natl Acad. Sci. USA* **108** 1240–5

Willmott C J and Matuura K 2001 Terrestrial air temperature and precipitation monthly and annual time series (1950–1999) *Centre for Climate Research* (http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html)

Wu D, Zhao X, Liang S, Zhou T, Huang K, Tang B and Zhao W 2015 Time-lag effects of global vegetation responses to climate change *Glob. Change Biol.* **21** 3520–31

Xie P and Arkin P A 1997 Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs *Bull. Am. Meteorol. Soc.* **78** 2539–58

Xie P, Chen M, Yang S, Yatagai A, Hayasaka T, Fukushima Y and Liu C 2007 A gauge-based analysis of daily precipitation over East Asia *J. Hydrometeorol.* **8** 607–26

Xie Y, Wang X and Silander J A Jr 2015 Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts *Proc. Natl Acad. Sci. USA* **112** 13585–90

Zimmermann N E, Yoccoz N G, Edwards T C, Meier E S, Thuiller W, Guisan A, Schmatz D R and Pearman P B 2009 Climatic extremes improve predictions of spatial patterns of tree species *Proc. Natl Acad. Sci. USA* **106** 19723–8

Zhao M and Running S W 2010 Drought-induced reduction in global terrestrial net primary production from 2000 through 2009 *Science* **329** 940–3

Zhou L *et al* 2014 Widespread decline of Congo rainforest greenness in the past decade *Nature* **508** 86–90

Zhu Z *et al* 2016 Greening of the Earth and its drivers *Nat. Clim. Change* **6** 791–5

Zscheischler J *et al* 2014 Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data *Glob. Biogeochem. Cycles* **28** 585–600