

LETTER • **OPEN ACCESS**

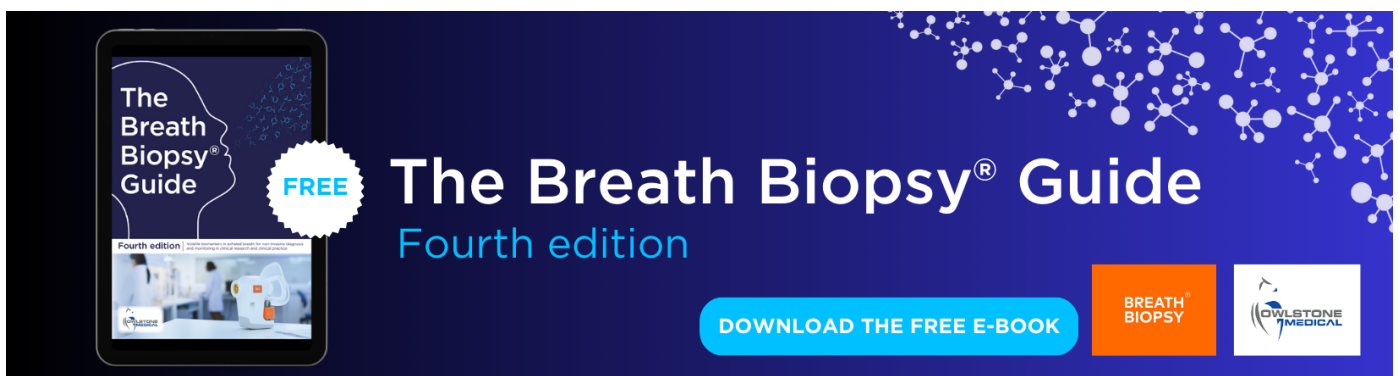
## Combining precipitation forecasts and vegetation health to predict fire risk at subseasonal timescale in the Amazon

To cite this article: Kátia Fernandes *et al* 2022 *Environ. Res. Lett.* **17** 074009

View the [article online](#) for updates and enhancements.

You may also like

- [Design and Implementation of BIM-based Fire Risk Assessment System](#)  
Hui Zhang
- [Assessing decadal variability of subseasonal forecasts of opportunity using explainable AI](#)  
Marybeth C Arcodia, Elizabeth A Barnes, Kirsten J Mayer et al.
- [Multi-stakeholder analysis of fire risk reduction in a densely populated area in the Netherlands: a case-study in the Veluwe area](#)  
Evelien de Hoop, Hilde J H Brouwers, Sophie L Buijs et al.



The Breath Biopsy® Guide  
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

## Combining precipitation forecasts and vegetation health to predict fire risk at subseasonal timescale in the Amazon

## OPEN ACCESS

RECEIVED  
30 December 2021REVISED  
1 June 2022ACCEPTED FOR PUBLICATION  
8 June 2022PUBLISHED  
20 June 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Kátia Fernandes<sup>1,2,\*</sup> , Michael Bell<sup>1</sup>  and Ángel G Muñoz<sup>2</sup> <sup>1</sup> Department of Geosciences, University of Arkansas, 340 N Campus Dr, Fayetteville, AR, United States of America<sup>2</sup> International Research Institute for Climate and Society, Columbia University, 61 Route 9W, Palisades, NY 10964, United States of America

\* Author to whom any correspondence should be addressed.

E-mail: [kdfernan@uark.edu](mailto:kdfernan@uark.edu)**Keywords:** Amazon, subseasonal forecast, fire riskSupplementary material for this article is available [online](#)**Abstract**

Current forecast systems provide reliable deterministic forecasts at the scale of weather (1–7 days) and probabilistic outcomes at the scale of seasons (1–9 months). Only in recent years research has begun transitioning to operational settings to provide numerical predictions for a lead time of 2–4 weeks, a timescale known as subseasonal. The Subseasonal Experiment (SubX) multi-model ensemble mean precipitation forecast (2017–2021) for days 8–14 (week-2 forecast) is used as a covariate in logistic regression models to predict fire risk in the Amazon. In a complementary experiment, a vegetation health index (VHI) is added to SubX precipitation forecasts as a predictor of fires. We find that fire risk can be skillfully assessed in most of the Amazon where fires occur regularly. In some sectors, SubX week-2 precipitation alone is a reliable predictor of fire risk, but the addition of VHI as a predictor results both in (a) a larger portion of the Amazon domain with skillful forecasts and; (b) higher skill in some sectors. By comparing two sectors of the Amazon, we find that the added information provided by VHI is most relevant where the mosaic of land covers includes savannas and grassland, whereas SubX precipitation can be used as the sole predictor for week-2 fire risk forecast in areas where the mosaic of land cover is dominated by forests. Our results illustrate the potential for using numerical model forecasts, at the subseasonal timescale, in combination with satellite remote sensing of vegetation to obtain skillful fire risk forecasts in the Amazon. The operationalization of the methods presented in this study could allow for better preparedness and fire risk reduction in the Amazon with a lead time greater than a week.

**1. Introduction**

The use of statistical methods to transform numerical model outputs into calibrated seasonal predictions is a well-established field in climate research with applications in agriculture (Esquivel *et al* 2018, Shelia *et al* 2019, Fernandes *et al* 2020, Pons *et al* 2021), water resources (Muñoz *et al* 2010, Robertson *et al* 2014, Najafi *et al* 2021) and public health (Muñoz *et al* 2017, 2020, Landman *et al* 2020, Colón-González *et al* 2021). In ecosystems management, one type of application consists of combining statistical methods with numerical predictions to assess or forecast seasonal fires (Aragão *et al* 2007, Murdiyarto

and Adiningsih 2007, van der Werf *et al* 2008, Field *et al* 2009, Fernandes *et al* 2017). Sea surface temperature (SST) in the Atlantic and the Pacific basins are good indicators of anomalous precipitation, at various timescales, in the Amazon (Fu *et al* 2001, Espinoza *et al* 2013, Fernandes *et al* 2015, Marengo and Espinoza 2016). Thus, SST forecast and observations can be used to predict fire occurrence in the Amazon (Chen *et al* 2011, Fernandes *et al* 2011) and other fire prone regions (Ceccato *et al* 2014, Chen *et al* 2016, Turco *et al* 2018) with a lead time of 1–4 months. At the other end of the prediction timescale lies fire weather forecasts (1–7 days), varying in spatial coverage from global

(Field *et al* 2015, Di Giuseppe *et al* 2020) to regional, where models are calibrated to better represent the environmental characteristics of the region (IDEAM 2015, Aliaga Nestares *et al* 2018, Setzer *et al* 2019, Martins *et al* 2020).

While both timescales provide decision-makers with much needed information, seasonal forecasts can be too long a horizon for the positioning of resources and the fire weather can be too short. In between lies the 2-to-4 week subseasonal timescale, for which numerical prediction has recently been established by research-to-operation efforts such as the S2S Prediction (Vitart *et al* 2017) and the Subseasonal Experiment (SubX) (Pegion *et al* 2019). We use the SubX model outputs, given the statistically significant skill of precipitation forecast over all months in South America and the facilitated data retrieval and manipulation offered by the International Research Institute for Climate and Society (IRI) Data Library (Blumenthal *et al* 2014).

We develop a hybrid (dynamical-statistical) model using the NextGen prediction system methodological approach (Munoz *et al* 2019, WMO 2020) to predict fire risk. NextGen involves designing, calibrating, building ensembles, and verifying objective forecasts by first identifying decision-relevant variables followed by the analysis of the physical mechanisms, sources of predictability and suitable candidate predictors. Here, the decision-relevant variable is fire risk, defined as the likelihood of active fires occurrence as a function of precipitation and vegetation health conditions. This definition might be considered 'event risk' in disaster or hazards literature (Sarewitz *et al* 2003), as it does not explicitly consider the vulnerability of specific communities to fire. The Visible Infrared Imaging Radiometer Suite (VIIRS) (Hillger *et al* 2013, Schroeder and Giglio 2017) active fires are modeled as a function of the SubX multimodel ensemble mean (MME) week-2 precipitation forecast, for the 2017–2021 period. A complementary experiment includes the satellite-based NOAA Vegetation Health Index (VHI, Kogan 1997, 2001) as an additional predictor. VHI is updated in real-time and it reliably detects ecosystem sensitivity to meteorological conditions at various timescales (Andujar *et al* 2017, Kogan *et al* 2017). We then take a closer look at the skill of the statistical models in two sectors of the Amazon, each representative of a distinct fire season. Lastly, we explore the role of contrasting land covers on fire risk models' performance in each of these two sectors.

## 2. Data and methods

### 2.1. VIIRS active fires (VNP14IMG)

Fire occurrence is estimated using the active fire product from the VIIRS instrument on board of the Suomi National Polar-orbiting Partnership satellite. The daily VIIRS product at a resolution of 375 m

(VNP14IMG) consists of a fire mask with ten classes. Three identify low, nominal and high confidence levels of active fire detection while the other seven are classes unrelated to fires, such as water and clouds (table 1 of Schroeder and Giglio 2017). We resampled all variables to an intermediary grid of 0.22° latitude-longitude resolution within which, every active fire of nominal and high confidence is counted. The resampled daily active fires are then averaged to other timescales as needed (sections 2.5.1 and 2.5.2). Although the focus of our study is the Amazon biome, the full domain of the analyses comprises of tropical South America (81.25° W–44.25° W and 19.5° S–11.5° N). The VIIRS data is available from January 2012 to current.

### 2.2. Subseasonal precipitation forecast (SubX)

SubX (Kirtman *et al* 2017, Pegion *et al* 2019) consists of a multimodel publicly available database of historical re-forecasts and real-time forecasts. The historical re-forecasts start in 1999 and end between 2014 and 2016 depending on the model. Real-time forecasts are available from six groups and include forecasts for at least 32 days beyond the initialization date (table 1). The real-time forecasts start at various times in 2017. By the 5 August initialization date, five of the six models produce forecasts except for NCEP-CFSv2 which starts on 15 October. All models maintain up-to-date forecasts on a weekly (or more frequent) basis that can be accessed through IRI Data Library (Blumenthal *et al* 2014). Although combined historical re-forecast and real-time forecast provide data for over 20 years, we use only output from the real-time forecasts, beginning in 2017, as the number of ensemble members in each model differs from historical re-forecast to real-time periods and a multi-model mean calculated over 1999–2021 would be inconsistent. The SubX models' spatial resolution of 1° latitude-longitude, was linearly interpolated to 0.22° latitude-longitude grid to match the resolution of aggregated VIIRS data (section 2.1). Spatially downscaling numerical model outputs can introduce noise in interpolated precipitation (Wood *et al* 2004), especially during the wet season. This should be less of an issue during the dry season. To test this assumption, we compared the distribution of a sample consisted of SubXPr multimodel mean week-2 forecast at its original resolution (1° latitude-longitude) with the linearly downscaled resolution (0.22° latitude-longitude) sample during one dry season in southern Amazon. Both samples include every gridcell within the domain of coordinates 76° W–57° W and 13° S–4° S and the 13 weeks of the 2018 dry season (July–September). Using a two-sample Kolmogorov–Smirnov test, we tested the null hypothesis that both samples came from the same continuous distribution. The test result does not reject the null hypothesis at the 1% significant level.

**Table 1.** SubX model names, number of real-time forecast ensemble members, length of forecast and corresponding references.

Model	Institution	# of members	Forecast length in days (1st initialization date in 2017)	References
NCEP-CFSv2	National Centers for Environmental Prediction (NCEP)	4	45 (15 October)	Saha <i>et al</i> (2014)
EMC-GEFS	NCEP Environmental Modeling Center	21	35 (30 June)	Zhu <i>et al</i> (2014), Zhou <i>et al</i> (2016), Zhou <i>et al</i> (2017)
GMAO-GEOS	National Aeronautics and Space Administration Global Modeling and Assimilation Office	4	45 (25 July)	Rienecker <i>et al</i> (2008), Koster <i>et al</i> (2000), Molod <i>et al</i> (2012), Reichle and Liu (2014), Field <i>et al</i> (2015)
RSMAS-CCSM4	University of Miami Rosenstiel School of Marine and Atmospheric Science (running National Center for Atmospheric Research CCSM4)	9	45 (25 June)	Infanti and Kirtman (2016)
NRL-NESM	Naval Research Laboratory	1	45 (5 August)	Metzger <i>et al</i> (2014), Hogan <i>et al</i> (2014)
ESRL-FIM	National Oceanic and Atmospheric Administration Earth System Research Laboratory	4	32 (2 August)	Sun <i>et al</i> (2018a), (2018b)

The SubX real-time precipitation forecast (SubXPr) is arranged to reproduce 52 week-2 (days 8–14) forecast period per year by:

- Calculating, for each SubX model, the ensemble members mean week-2 precipitation forecast corresponding to 52 annual non-overlapping weeks, and a final sample of 208 elements per model (52 week-2 forecasts per year times 4 years. Mid August 2017–mid August 2021). See table S1 for details.
- The MME mean consist of week-2 forecasts averaged among the six SubX models.

Constructing a multi-model mean of week-2 real-time forecasts is a challenging task. The different models are initialized on different days of the week and at different intervals. To take advantage of Wednesday initializations in three of the six models, we fix non-overlapping week-2 real-time forecasts (average of forecast days 8–14) from Thursday (day 8) to Wednesday (day 14). For the other three models, week-2 real-time forecast varies from average forecast days 9–15 to average forecast days 12–18 depending on the SubX model and the forecast week (table 2).

### 2.3. Vegetation health index (VHI)

Fire ignition in the Amazon results from human activities, but the intensity and spread can be

regulated by natural processes, such as vegetation water stress (Asner and Alencar 2010). One indicator of vegetation stress is the satellite-based NOAA-STAR VHI calculated from the vegetation condition index (VCI), a proxy for vegetation moisture, and temperature condition index (TCI), a proxy for thermal effects (Kogan 2001)

$$VCI = 100 \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \quad (1)$$

$$TCI = 100 \frac{T_{\max} - T}{T_{\max} - T_{\min}} \quad (2)$$

$$VHI = \alpha VCI + (1 + \alpha) TCI, \quad \text{where } \alpha = 0.5. \quad (3)$$

VHI is available at 4 km spatial resolution, which was also re-gridded to the 0.22° latitude-longitude resolution. Near real-time weekly-averaged VHI is released on the day of the week corresponding to 8 January of each year. In 2018, for example, the product is updated every Monday (table 2) for a total of 52 global maps per year. For VHI to be used as a predictor, the data needs to be available by the weekly initialization date of the SubX multi-model mean precipitation forecast, and this varies within the years of study. From 2012 to current, the indices are calculated using the VIIRS instrument (Kogan *et al* 2015, 2017).

**Table 2.** Example of initialization dates used to produce a multi-model mean week-2 forecast for target date 18–24 January 2018, shaded blue in the table. The green cells mark initialization dates for each model. The X marks the Wednesday (10 January) initialization for a week-2 (average of forecast days 8–14) forecast and the closest prior dates for models with initializations on days other than Wednesday. The actual lead time for each forecast is shown below the target week-2 forecast period. The dates of observed VHI and VIIRS active fires corresponding to the example SubX precipitation target week are also marked in the table.

	Week day	F	S	S	M	T	W	T	F	S	S	M	T	W	Week-2 (Th-Wed) forecast						
	Initialization day	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NCEP-CFSv2	Dly						X								8	9	10	11	12	13	14
EMC-GEFS	W							X							8	9	10	11	12	13	14
GMAO-GEOS	5 days		X												12	13	14	15	16	17	18
RSMAS-CCSM4	Su			X											11	12	13	14	15	16	17
ESRL-FIM	W							X							8	9	10	11	12	13	14
NRL-NESM	Sa, Su, Mo, Tu						X								9	10	11	12	13	14	15
VHI	Mo				X																
VIIRS-Weekly AF																					

The 2017–2021 VHI is obtained for dates corresponding to the available SubX week-2 precipitation forecasts.

**2.4. Land cover-Mapbiomas**

We use Mapbiomas Land Cover Collection-6 Brazil (Souza et al 2020) in two areas of the Amazon chosen for a closer evaluation of fire risk model skills. We chose the year 2019 land-cover as representative of the 2017–2021 period.

**2.5. Methods**

**2.5.1. Active fire season**

The analysis relating precipitation, fires and vegetation health is conducted for the peak fire season. The nearly 10 years of VIIRS data (January 2012–mid-August 2021) is used to determine the climatological most active fire trimester at each gridcell by calculating monthly averages of daily VIIRS active fire data (section 2.1) and then calculating the climatology for the annual 12 overlapping trimesters (JFM, FMA, etc). The trimester with highest fire activity (FA) is then identified (figure 1).

**2.5.2. Weekly data**

Once the most active fire trimester is identified, the VIIRS, SubXPr and VHI data for the 13 weeks within that trimester are fitted in a logistic regression model. If a gridcell’s most active fire trimester is September–November, data for the season’s 13 weeks of years 2017, 2018, 2019 and 2020 (total of 52 weeks = 13 weeks times 4 years) is fitted in the statistical model. For March–May, for example, the data sample contains the trimester’s 13 weeks of 2018, 2019, 2020 and 2021 given the SubX multi-model forecast is available from August 2017 to August 2021.

**2.5.3. Statistical analysis of fire probability**

Fires are used in the Amazon as a tool to clear debris from deforestation and to manage agropastoral land.

They can occur concentrated in a few weeks as burning is planned for times when vegetation is dry. Thus, the timeseries of weekly FA may show a mix of high activity interspaced by no fires. In the majority of our study domain, weeks with no fires occur at least during 20% of the time in the dry season (figure S1) and a two-mode behavior (fire/no-fire) is then fitted to predictors using logistic regression.

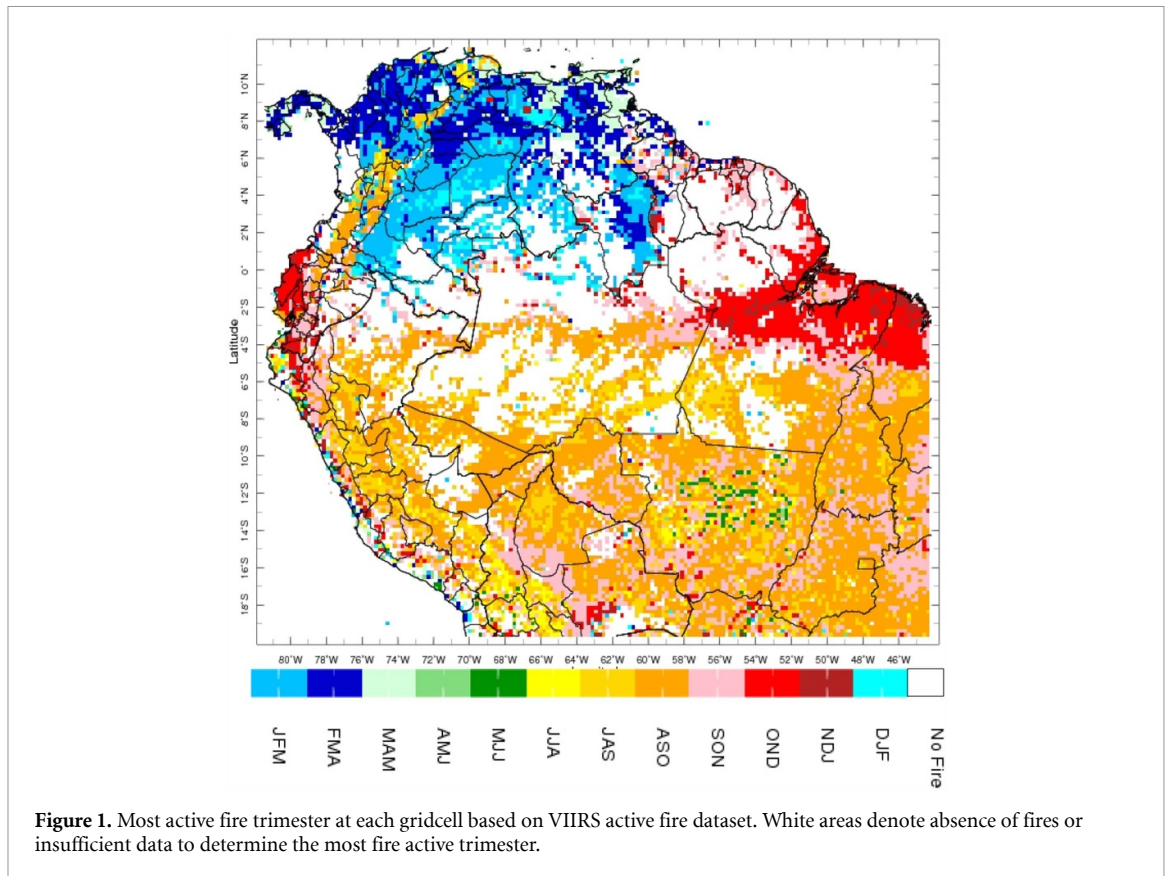
To assess the potential for predicting weekly FA, we choose logistic regression expressed as:

$$\ln \left( \frac{P(FA = 1)}{1 - P(FA = 1)} \right) = \alpha + \beta_1 X_1 + \dots + \beta_n X_n \quad (4)$$

where  $P(FA = 1)$  is the probability of fire occurrence (FA = 1) modeled as a linear combination of predictors  $X_n$  with slopes  $\beta_n$  and an intercept  $\alpha$ . The probability of fire occurrence is modeled as a function of non-overlapping week-2 SubXPr and VHI. The FA binary response is coded 1 for values greater than 0 and 0 for no fire within a 0.22° latitude-longitude gridcell.

The logistic regression models were fit using Matlab 2020a fitglm function (MathWorks 2020) for a binomial distribution. In one experiment, FA was fit to SubXPr and in the other FA was fit to SubXPr and VHI. The experiments are henceforth called FA\_f(SubXPr) and FA\_f(SubXPr&VHI), respectively. The logistic regression models were cross-validated using the leave-one-out option, which corresponds to leaving out one week of paired datasets at each iteration. The assessment of gridcell’s logistic regression model fit is evaluated by determining what portion of observed positives is classified as positive (sensitivity) and what portion of observed negatives is classified as negative (specificity). A plot of sensitivity against 1-specificity describes the receiver operating characteristics (ROC, Wilks 2011) and goodness of fit can be assessed as the area under the ROC





curve (AUC). An AUC of 0.5 represents discrimination expected by chance to predict positive outcomes. AUC above 0.5 mean that the positive outcome has a higher predicted probability than the negative subject, and the logistic regression model is considered skillful. A value of 1 stands for a perfect model.

### 3. Results

#### 3.1. SubXPr and VHI as predictors of fire probability

We discuss our results for a subset of gridcells with intermediate and high values of AUC ( $>0.6$ ). In some sectors, using SubXPr week-2 forecast as the sole predictor (FA\_f(SubXPr) results in AUC  $> 0.7$ , notably in the southwestern Brazilian Amazon marked with a box in figure 2(a) and in eastern equatorial Amazon. AUC values show improvement for experiment FA\_f(SubXPr&VHI), especially in the southern Amazon where an extensive area shows AUC  $> 0.6$  (figure 2(b)). An exception is the eastern equatorial Amazon, where precipitation alone results in better performance both in absolute values and in number of gridcells with high AUC.

The results in figure 2 illustrate the potential for relying on SubX multi-model mean precipitation forecasts to determine the probability of fire occurrence in the Amazon with a lead of 8–14 days (week-2). Moreover, by including the state of vegetation as an indicator of fire risk, the logistic regression models

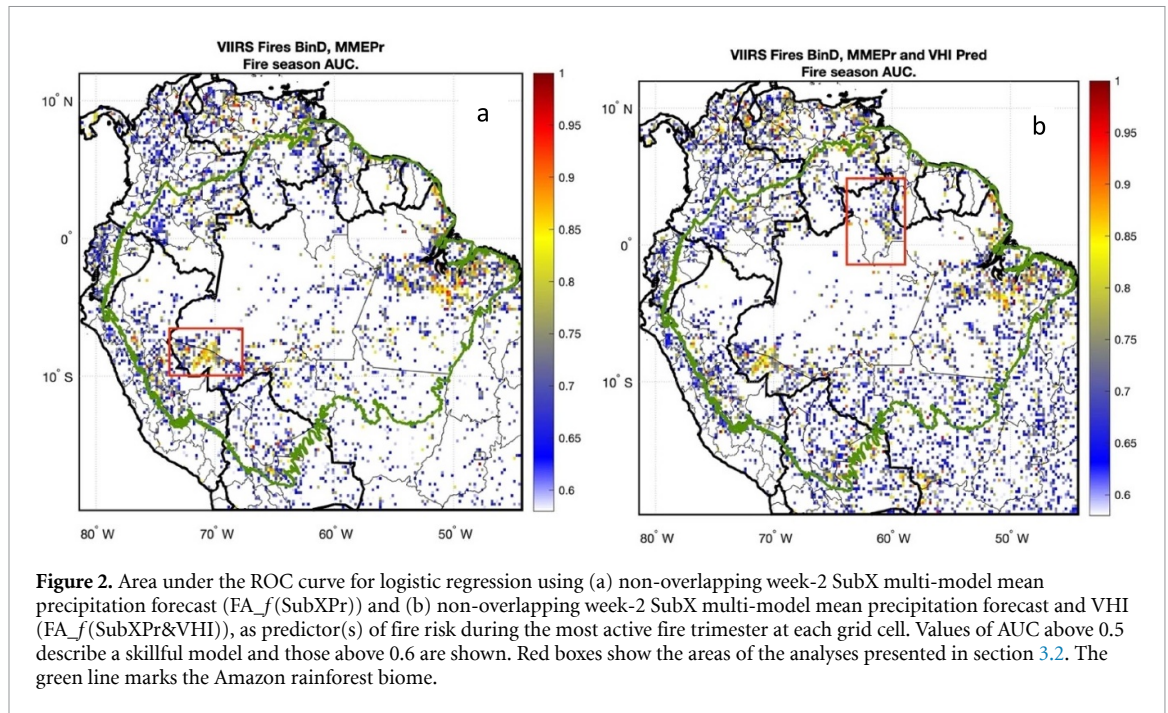
perform better over a larger sector of tropical South America.

#### 3.2. Model performance in southern and northern Amazon

To further explore fire occurrence response to SubXPr and VHI, we selected two sectors of the Amazon with a cluster of high AUCs for each of the experiments. In the southern Amazon (73.6° W–68.5° W, 9.8° S–6.5° S), where the main climatological fire season is August–October, high AUC is found for experiment FA\_f(SubXPr) (figure 2(a)). In the northern sector (63° W–59.2° W, 1.7° S–4.8° N), where the climatological fire season is January–April, a cluster of high AUC is found for experiment FA\_f(SubXPr&VHI) (figure 2(b)).

##### 3.2.1. Forecast time series

We calculate, for each spatial domain, how many gridcells predicted the correct outcome among those with AUC  $> 0.6$ . A successful fire forecast means fire probability higher than 0.5 in a gridcell that detected fire or fire probability below or equal 0.5 in a gridcell that detected no-fire. The gridcells with a ‘hit’ are counted and a ratio between the number of correct hits and the total number of gridcells with AUC  $> 0.6$  within the domain is calculated for every forecast week (figure 3). Given the first SubXPr forecast is available for the week of 17–23 August 2017, the weeks used in the cross-validated models within the southern domain (fire season August–October) start in



2017 and end in 2021. This corresponds to a total of 52 weeks distributed in 11 weeks in 2017, 13 weeks in 2018, 2019 and 2020, and 2 weeks in 2021. The weeks used for the northern domain (fire season January–April) start in 2018 and end in 2021 (table S1).

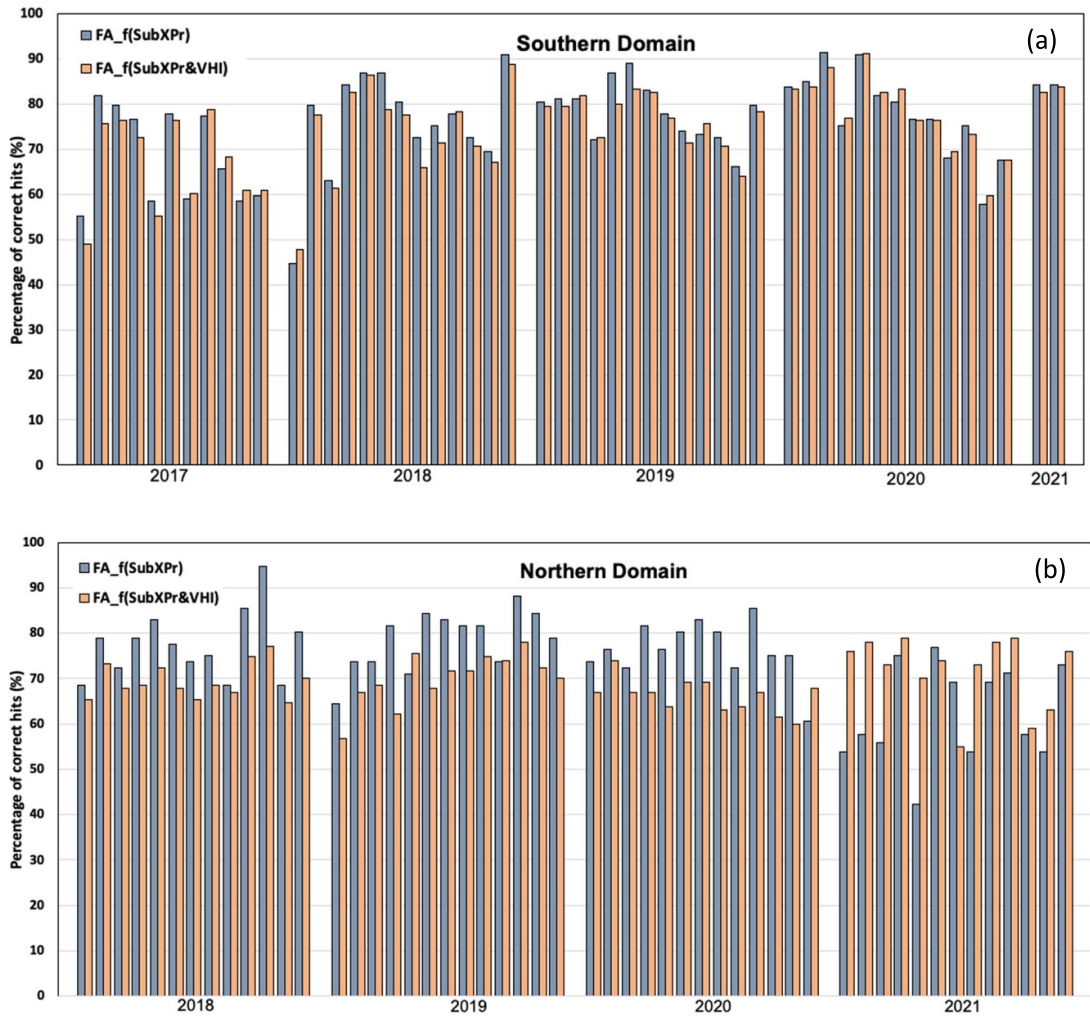
Within the southern domain the number of gridcells with AUC > 0.6 are 154 and 161 for FA<sub>f</sub>(SubXPr) and FA<sub>f</sub>(SubXPr&VHI), respectively. This similar number of skillful models can be observed in figures 2(a) and (b), where the gridcells with AUC > 0.6 are shaded. Collectively, both experiments result in similar performance with an average of 75.6% of gridcells correctly predicting fire risk in experiment FA<sub>f</sub>(SubXPr) and 74.3% in the FA<sub>f</sub>(SubXPr&VHI) experiment (figure 3(a)).

In the northern domain the number of gridcells with AUC > 0.6 are 76 and 127 for experiments FA<sub>f</sub>(SubXPr) and FA<sub>f</sub>(SubXPr&VHI), respectively. On average, 74% of gridcells correctly forecast outcomes for the FA<sub>f</sub>(SubXPr) experiment and 69% for FA<sub>f</sub>(SubXPr&VHI) (figure 3(b)). Relatively, the models in experiment FA<sub>f</sub>(SubXPr) outperform those of experiment FA<sub>f</sub>(SubXPr&VHI), but the latter represents a considerably larger overall area—more gridcells—with skillful forecasts. The average 74% correct hits from FA<sub>f</sub>(SubXPr) corresponds to 56 gridcells, whereas the 69% from FA<sub>f</sub>(SubXPr&VHI) corresponds to 87 gridcells, indicating that the addition of VHI as predictor provides a more useful set of information in the northern domain, especially for extreme VHI values (see figure 4 discussion). In grasslands, common in the northern domain (section 3.2.3), fires occur regularly, and fire risk will be mainly determined by whether it rains or not. However, if at the start of

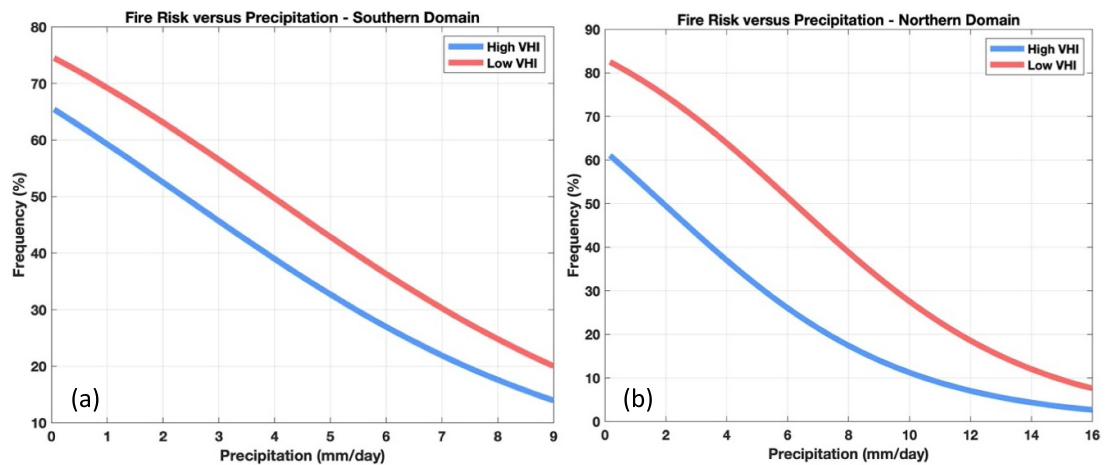
and during the dry season VHI is ‘neutral to high-wetness’, VHI becomes a valuable indicator of fire-risk as absence of rain will not immediately translate in higher fire-risk given the high-wetness state of vegetation. The high VHI state in 2021 (figure S2) would explain the models’ performance in that year (figure 3(b)), when SubXPr alone did a poorer job in predicting fire risk than SubXPr and VHI together.

### 3.2.2. Fire risk sensitivity to SubXPr and VHI

To test fire risk sensitivity to SubXPr and VHI in two domains identified in section 3.2, we fitted a logistic regression model to one continuous sample comprised of data from gridcells within each domain and all 52 weeks. Our objective is to evaluate if one single logistic regression model fitted to the aggregated data from each domain exhibits characteristics similar to the gridcells’ individual models, namely low sensitivity to VHI in the southern domain and high sensitivity in the northern domain. The southern domain has the spatial dimension of 25 (longitude) per 17 (latitude) cells for a total of 425 at each timestep. A sample derived from this domain contains a maximum of 22 100 elements (425 gridcells times 52 weeks). In the northern domain the spatial dimension is 18 (longitude) per 32 (latitude) resulting in 576 gridcells at each timestep for a maximum of 29 952 elements (576 gridcells times 52 weeks). The sample sizes for the southern and northern sectors contain in fact 13 185 and 14 201 elements, respectively, once all the gridcells with an undefined fire season (figure 1) and occasional missing data are removed. We then reproduce the experiment FA<sub>f</sub>(SubXPr&VHI) using the aggregate covariates for each domain and present



**Figure 3.** Ratio between the number of gridcells with correct week-2 fire risk forecast and the total number of gridcells for which a skillfull model exist ( $AUC > 0.6$ ) within the southern (a) and northern (b) domains. A correct outcome means the logistical model forecasted a probability greater than 0.5 for fire occurrence in gridcells where fires were detected and lower or equal to 0.5 for gridcells with no-fire.



**Figure 4.** Fire probability, measured as probability of occurrence of fires (FA), as a function of precipitation for low (red) and high (blue) VHI conditions within the southern (a) and northern (b) domains.



**Table 3.** Logistic regression coefficients ( $\beta$ ), standardized errors (SE), statistical significance ( $P$ ), and area under the receiver operating characteristics (ROC) curve (AUC) for regression predictors of fire risk derived from aggregated gridcells within the southern and northern domains.

Southern domain	$\beta$	SE	$P$	AUC
Coefficients				0.63
Intercept	1.3	0.067	<0.01	
SubXPr	-0.27	0.012	<0.01	
VHI	-0.01	0.002	<0.01	
Northern domain				0.73
Intercept	1.99	0.064	<0.01	
SubXPr	-0.25	0.007	<0.01	
VHI	-0.021	0.001	<0.01	

the logistic regression models' coefficients, standard error, statistical significance, and AUC in table 3.

The SubXPr and fire relationship sensitivity to VHI is evaluated from setting two extreme VHI scenarios. Low VHI is the average of the 10% lowest values in the aggregated time series used to fit the statistical models and high VHI is the average of the 10% highest values. This is done for both domains using equations (5) and (6). Low VHI values are 21 and 19, and high VHI are 62 and 72 for the southern and northern domains respectively

$$\text{Southern Domain } P(AF = 1) = \frac{e^{1.3 - 0.27 * \text{SubXPr} - 0.01 * \text{VHI}}}{1 + e^{1.3 - 0.27 * \text{SubXPr} - 0.01 * \text{VHI}}} \quad (5)$$

$$\text{Northern Domain } P(AF = 1) = \frac{e^{1.99 - 0.25 * \text{SubXPr} - 0.021 * \text{VHI}}}{1 + e^{1.99 - 0.25 * \text{SubXPr} - 0.021 * \text{VHI}}} \quad (6)$$

Fire probability as a function of SubXPr for low VHI (vegetation stress) shows higher values in general than that describing fire probability for high VHI (no-stress/high-wetness) conditions in the southern domain (figure 4(a)). For precipitation of 3 mm d<sup>-1</sup> there is a 45% chance of observing fires under high VHI conditions, increasing to 56% for low VHI, indicating low fire sensitivity to vegetation conditions regardless of precipitation amounts.

In the northern domain, a more robust shift of probabilities is observed, especially for lower values of precipitation (<7 mm d<sup>-1</sup>) as seen figure 4(b), describing a prominent role of vegetation health in determining fire risk. At 3 mm d<sup>-1</sup>, fire risk goes from 43% for high VHI to 69% for low VHI. Beyond 10 mm d<sup>-1</sup>, the two VHI scenarios converge to describe a dominant role of precipitation on fire risk regardless of the state of vegetation.

### 3.2.3. Land cover

These two sectors of the Amazon are characterized by different land covers as estimated by the Mapbiomas Land Cover classification for year 2019 (figure S3). We resampled the 30 meter resolution land cover map (Souza *et al* 2020) to the VHI 4 km original grid, by retaining the class with the highest frequency to

determine the land covers most influential for the VHI estimate. The southern domain contains 98% forest and nearly 2% pasture (figure 5(a)), whereas in the northern domain vegetation is more varied and comprised of 77% forest, 20% grassland and 1.6% pasture (figure 5(b)).

The vegetation distribution in the two sectors, provides a plausible explanation for the distinct fire risk sensitivity to VHI. In VHI, greenness normalized vegetation index (NDVI) is as a proxy for vegetation moisture conditions (equation (1)). However, reduced cloud cover and more light toward the end of the dry season can result in a green-up (higher NDVI) of the Amazon forest (Huete *et al* 2006). At the spatial scale of VHI (4 km), this green up could mask NDVI of other land covers within a gridcell, reducing the power of VHI as an indicator of fire risk in land covers that are more susceptible to burning but border green forest patches. Even if VHI can describe forest water stress, that does not necessarily mean susceptibility to fires. Understory humidity and proximity to fire sources, also play an important role in determining whether fires occur in forest land covers.

In contrast, savannas and grasslands are more sensitive to short-term water stress and VHI responds to meteorological drought, or recovery from it, within a couple of weeks to a month (Sehgal *et al* 2021), better reflecting short-term vegetation sensitivity to fires (Anderson *et al* 2010, Zhao *et al* 2017) adding predictive power to fire risk assessment weeks in advance.

## 4. Discussion

While atmospheric scientists focus on the drivers of fires, fire ecologists focus on the atmospheric demand for moisture (e.g. vapor pressure deficit) and the potential for vegetation to become fuel (Fu *et al* 2021). Our study aims at reducing this gap by combining observed states of vegetation with precipitation forecasts. Using a SubX multimodel ensemble mean of week-2 precipitation forecasts, we find that fire probability can be skillfully predicted in large portions of the Amazon. By pairing week-2 SubXPr with VHI to predict fire probability, we find areas in which logistic models' skills improve upon those using only week-2 SubXPr forecast as a predictor. It is also evident that a greater number of gridcells in the domain of study present skillful statistical forecast for the experiment using both VHI and SubXPr as predictors of fire risk. This is especially true in the Brazilian arc of deforestation and in Bolivia that are characterized by land covers that are more susceptible to fires (grassland, fallow, and agropastoral land). In areas where land cover is dominated by forests, the role of vegetation health as a predictor of fire risk is modest. That is the case of the southern Brazilian Amazon, where fire risk is mainly driven by SubXPr and the probability of fire occurrence increases only about 10% for very

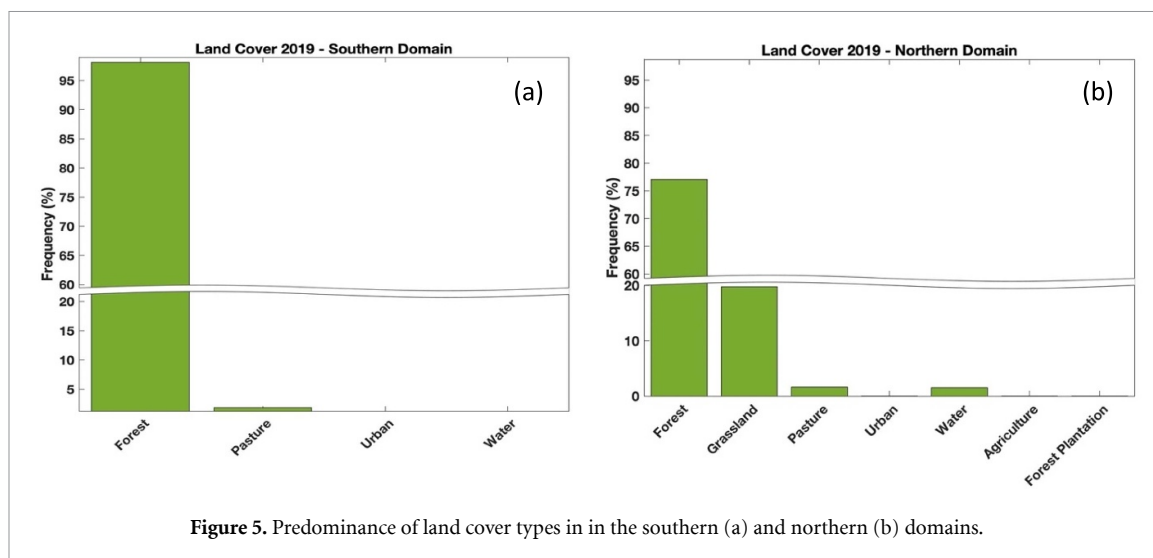


Figure 5. Predominance of land cover types in in the southern (a) and northern (b) domains.

low (water stressed) to very high VHI across different precipitation forecasts.

In contrast, in the northern Amazon where savannas and grasslands are more prevalent, VHI status helps determine fire risk in the weeks ahead. This is seen in the larger number of gridcell with skilfull fire risk forecast for experiment FA<sub>f</sub>(SubXPr&VHI) compared to FA<sub>f</sub>(SubXPr). The sensitivity test also indicates that fire risk increases by as much as 28% from high VHI (non-stressed) to low VHI for lower precipitation values indicating the more prominent role of vegetation status in determining fire risk.

## 5. Conclusions

Our study shows the potential for using precipitation subseasonal forecasts in combination with vegetation health in a hybrid dynamical-statistical model to forecast fire risk. The operationalization of the methods presented in this study have the potential to inform better practices for fire prevention and mitigation in the Amazon with lead time beyond one week, which is a crucial timescale for fire management resources allocation. Continuous improvement of statistical models' accuracy can potentially be achieved by using satellite vegetation conditions at higher spatial resolution, which would allow for better discrimination between forests and non-forest land covers, and by including more years of SubX precipitation forecast.

## Data availability statement

VNP14IMG (VIIRS Fires)- <https://firms.modaps.eosdis.nasa.gov/download/>

VHI: [www.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/Blended\\_VH\\_4km/geo\\_TIFF](http://www.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/Blended_VH_4km/geo_TIFF)

Mapbiomas: <http://mapbiomas.org>

The data that support the findings of this study are openly available at the following URL/DOI: [10.7916/D8PG249H](https://doi.org/10.7916/D8PG249H).

## Acknowledgments

We acknowledge the agencies that support the SubX Project and the modeling groups for producing and making available their model output, including the National Centers for Environmental Prediction (NCEP), the NCEP Environmental Modeling Center, the National Aeronautics and Space Administration Global Modeling and Assimilation Office, the National Center for Atmospheric Research and University of Miami Rosenstiel School of Marine and Atmospheric Science, the Naval Research Laboratory, the National Oceanic and Atmospheric Administration Earth System Research Laboratory, and Environment and Climate Change Canada. We wish to thank Columbia University's International Research Institute for Climate and Society (IRI) Climate Data Library for disseminating the SubX data (<https://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/>; DOI: [10.7916/D8PG249H](https://doi.org/10.7916/D8PG249H)).

K F and M B, wish to acknowledge the support of the SERVIR-Amazonia, a joint development initiative of the National Aeronautics and Space Administration (NASA) and the United States Agency for International Development (USAID). A G M was partially supported by NOAA Grant: NA18OAR4310275 and ACToday, the first Columbia World Project. We also thank the anonymous reviewers for the constructive comments and suggestions.

## Conflict of interest

The authors declare no competing interests.

## ORCID iDs

Kátia Fernandes  <https://orcid.org/0000-0001-8160-0769>

Michael Bell  <https://orcid.org/0000-0002-5867-9787>

Ángel G Muñoz  <https://orcid.org/0000-0002-2212-6654>

## References

- Aliaga Nestares V, Quispe Gutierrez N, Ramos Parado I and Rodriguez Zimmermann D 2018 Estudio de Condiciones Atmosféricas Favorables a los Incendios Forestales en el Perú (Lima) (available at: [www.senamhi.gob.pe/load/file/01401SENA-45.pdf](http://www.senamhi.gob.pe/load/file/01401SENA-45.pdf))
- Anderson L O et al 2010 Remote sensing detection of droughts in Amazonian forest canopies S2S *Wiley Online Libr.* **187** 733–50 (available at: <https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2010.03355.x>)
- Andujar E, Krakauer N Y, Yi C and Kogan F 2017 Ecosystem drought response timescales from thermal emission versus shortwave remote sensing *Adv. Meteorol.* **2017** 8434020
- Aragão L E O C, Malhi Y, Roman-Cuesta R M, Saatchi S, Anderson L O and Shimabukuro Y E 2007 Spatial patterns and fire response of recent Amazonian droughts *Geophys. Res. Lett.* **34** L07701
- Asner G P and Alencar A 2010 Drought impacts on the Amazon forest: the remote sensing perspective *New Phytol.* **187** 569–78
- Blumenthal M, Bell M, Del Corral J, Cousin R and Khomyakov I 2014 IRI data library: enhancing accessibility of climate knowledge *Earth Perspect.* **1** 19
- Ceccato P, Fernandes K, Ruiz D and Allis E 2014 Climate and environmental monitoring for decision making *Earth Perspect.* **1** 16
- Chen Y, Morton D C, Andela N, Giglio L and Randerson J T 2016 How much global burned area can be forecast on seasonal time scales using sea surface temperatures? *Environ. Res. Lett.* **11** 045001
- Chen Y, Randerson J T, Morton D C, DeFries R S, Collatz G J, Kasibhatla P S, Giglio L, Jin Y and Marlier M E 2011 Forecasting fire season severity in South America using sea surface temperature anomalies *Science* **334** 787–91
- Colón-González F J et al 2021 Probabilistic seasonal dengue forecasting in Vietnam: a modelling study using superensembles *PLoS Med.* **18** e1003542
- Di Giuseppe F, Vitolo C, Krzeminski B, Barnard C, MacIel P and San-Miguel J 2020 Fire weather index: the skill provided by the European Centre for medium-range weather forecasts ensemble prediction system *Nat. Hazards Earth Syst. Sci.* **20** 2365–78
- Espinoza J C, Ronchail J, Frappart F, Lavado W, Santini W and Guyot J L 2013 The major floods in the Amazonas River and Tributaries (Western Amazon Basin) during the 1970–2012 period: a focus on the 2012 flood *J. Hydrometeorol.* **14** 1000–8
- Esquivel A, Llanos-Herrera L, Agudelo D, Prager S D, Fernandes K, Rojas A, Valencia J J and Ramirez-Villegas J 2018 Predictability of seasonal precipitation across major crop growing areas in Colombia *Clim. Serv.* **12** 36–47
- Fernandes K et al 2011 North tropical Atlantic influence on western Amazon fire season variability *Geophys. Res. Lett.* **38** L12701
- Fernandes K, Giannini A, Verchot L, Baethgen W and Pinedo-Vasquez M 2015 Decadal covariability of Atlantic SSTs and western Amazon dry-season hydroclimate in observations and CMIP5 simulations *Wiley Online Libr.* **42** 6793–801
- Fernandes K, Muñoz A G, Ramirez-Villegas J, Agudelo D, Llanos-Herrera L, Esquivel A, Rodriguez-Espinoza J and Prager S D 2020 Improving seasonal precipitation forecasts for agriculture in the orinoquía Region of Colombia *Weather Forecast.* **35** 437–49
- Fernandes K, Verchot L, Baethgen W, Gutierrez-Velez V, Pinedo-Vasquez M and Martius C 2017 Heightened fire probability in Indonesia in non-drought conditions: the effect of increasing temperatures *Environ. Res. Lett.* **12** 054002
- Field R D et al 2015 Development of a global fire weather database *Nat. Hazards Earth Syst. Sci.* **15** 1407–23
- Field R D, van der Werf G R and Shen S S P 2009 Human amplification of drought-induced biomass burning in Indonesia since 1960 *Nat. Geosci.* **2** 185–8
- Fu R, Dickinson R E, Chen M and Wang H 2001 How do tropical sea surface temperatures influence the seasonal distribution of precipitation in the equatorial Amazon? *J. Clim.* **14** 4003–26
- Fu R, Hoell A, Mankin J, Sheffield A and Simpson I 2021 Tackling challenges of a drier, hotter, more fire-prone future *Eos* **102**
- Hillger D et al 2013 First-light imagery from Suomi NPP VIIRS *Bull. Am. Meteorol. Soc.* **94** 1019–29
- Hogan T et al 2014 The navy global environmental model *Oceanography* **27** 116–25
- Huete A R, Didan K, Shimabukuro Y E, Ratana P, Saleska S R, Hutyrá L R, Yang W, Nemani R R and Myneni R 2006 Amazon rainforests green-up with sunlight in dry season *Geophys. Res. Lett.* **33** L06405
- IDEAM 2015 *Protocolo para la Realización de Mapas de Zonificación de Riesgos a Incendios de la Cobertura Vegetal-Escala 1:100.000* (Bogota: Instituto de Hidrología, Meteorología y Estudios Ambientales) p 109
- Infanti J M and Kirtman B P 2016 Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America *J. Geophys. Res.* **121** 12690–701
- Kirtman B P et al 2017 The subseasonal experiment (SubX) *Chaos* **27**
- Kogan F N 1997 Global drought watch from space *Bull. Am. Meteorol. Soc.* **78** 621–36
- Kogan F N 2001 Operational space technology for global vegetation assessment *Bull. Am. Meteorol. Soc.* **82** 1949–64
- Kogan F, Guo W and Ding H 2015 NOAA JPSS Visible Infrared Imaging Radiometer Suite (VIIRS) Vegetation Health and Drought Products (VHDP) from NDE NOAA National Centers for Environmental Information (<https://doi.org/10.7289/V58W3BCR>)
- Kogan F, Guo W and Yang W 2017 SNPP/VIIRS vegetation health to assess 500 California drought *Geomat. Nat. Hazards Risk* **8** 1383–95
- Koster R D, Suarez M J, Ducharne A, Stieglitz M and Kumar P 2000 A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure *J. Geophys. Res. Atmos.* **105** 24809–22
- Landman W A, Sweijd N, Masedi N and Minakawa N 2020 The development and prudent application of climate-based forecasts of seasonal malaria in the Limpopo province in South Africa *Environ. Dev.* **35** 100522
- Marengo J A and Espinoza J C 2016 Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts *Int. J. Climatol.* **36** 1033–50
- Martins G, Nogueira J, Setzer A and Morelli F 2020 Comparison between different versions of INPE'S fire risk model for the Brazilian Biomes 2020 *IEEE Latin American GRSS and ISPRS Remote Sensing Conf. LAGIRS 2020—Proc.* (Institute of Electrical and Electronics Engineers Inc.) pp 380–5
- MathWorks, Inc. 2020 MATLAB. Version 2020a (available at: [www.mathworks.com/](http://www.mathworks.com/))
- Metzger J E et al 2014 UU navy operational global ocean and arctic ice prediction systems *Oceanography* **27** 32–43
- Molod A, Takacs L, Suarez M, Bacmeister J, Song I-S and Eichmann A 2012 The GEOS-5 atmospheric general circulation model: mean climate and development from

- MERRA to fortuna NASA/TM-2012-104606-Vol-28 (National Aeronautics and Space Administration) (available at: <https://ntrs.nasa.gov/citations/20120011790>)
- Muñoz Á G et al 2010 An environmental watch system for the Andean countries: el observatorio andino *Bull. Am. Meteorol. Soc.* **91** 1645–52
- Muñoz A G, Thomson M C, Stewart-Ibarra A M, Vecchi G A, Chourio X, Nájera P, Moran Z and Yang X 2017 Could the recent zika epidemic have been predicted? *Front. Microbiol.* **8** 1291
- Munoz A, Robertson A W, Mason S J, Goddard L M, Pons D, Turkington T, Acharya N, Siebert A and Chourio X M 2019 NextGen: a next-generation system for calibrating, ensembling and verifying regional seasonal and subseasonal forecasts *American Geophysical Union, Fall Meeting 2019 A23U-3024* (available at: <https://ui.adsabs.harvard.edu/abs/2019AGUFM.A23U3024M/abstract>)
- Muñoz G, Chourio X, Rivière-Cinnamon A, Diuk-Wasser M A, Kache P A, Mordecai E A, Harrington L and Thomson M C 2020 AeDES: a next-generation monitoring and forecasting system for environmental suitability of Aedes-borne disease transmission *Sci. Rep.* **10** 12640
- Murdiyarso D and Adiningsih E S 2007 Climate anomalies, Indonesian vegetation fires and terrestrial carbon emissions *Mitig. Adapt. Strateg. Glob. Change* **12** 101–12
- Najafi H, Robertson A W, Massah Bavani A R, Irannejad P, Wanders N and Wood E F 2021 Improved multi-model ensemble forecasts of Iran's precipitation and temperature using a hybrid dynamical-statistical approach during fall and winter seasons *Int. J. Climatol.* **41** 5698–725
- Pegion K et al 2019 The subseasonal experiment (SubX): a multimodel subseasonal prediction experiment *Bull. Am. Meteorol. Soc.* **100** 2043–60
- Pons D, Muñoz Á G, Meléndez L M, Chocooj M, Gómez R, Chourio X and González Romero C 2021 A coffee yield next-generation forecast system for rain-fed plantations: the case of the Samalá Watershed in Guatemala *Weather Forecast.* **36** 2021–38
- Reichle R H and Liu Q 2014 *Observation-corrected precipitation estimates in GEOS-5* NASA/TM-2014-104606 The National Aeronautics and Space Administration (available at: <https://ntrs.nasa.gov/citations/20150000725>)
- Rienecker M M et al 2008 The GEOS-5 data assimilation system—documentation of versions 5.0.1 and 5.1.0, and 5.2.0 vol 27 NASA/TM-2008-104606-VOL-27 (The National Aeronautics and Space Administration) p 92 (available at: <https://ntrs.nasa.gov/api/citations/20120011955/downloads/20120011955.pdf?attachment=true>)
- Robertson A W, Baethgen W, Block P, Lall U, Sankarasubramanian A, de Assis de Souza Filho F and J Verbist K M 2014 Climate risk management for water in semi-arid regions *Earth Perspect.* **1** 12
- Saha S et al 2014 The NCEP climate forecast system version 2 *J. Clim.* **27** 2185–208
- Sarewitz D, Pielke R and Keykhah M 2003 Vulnerability and risk: some thoughts from a political and policy perspective *Risk Anal.* **23** 805–10
- Schroeder W and Giglio L 2017 Visible infrared imaging radiometer suite (VIIRS) 375 m & 750 m active fire detection data sets based on NASA VIIRS land science investigator processing system {(SIPS)} reprocessed data-Version 1 product User's guide Version 1.2 (available at: [https://lpdaac.usgs.gov/documents/132/VNP14\\_User\\_Guide\\_v1.3.pdf](https://lpdaac.usgs.gov/documents/132/VNP14_User_Guide_v1.3.pdf))
- Sehgal V, Gaur N and Mohanty B P 2021 Global flash drought monitoring using surface soil moisture *Water Resour. Res.* **57** 1–25
- Setzer A, Sismanoglu R and Martins G 2019 *Métododo Cálculo do Risco de Fogo do Programa do INPE* (São José dos Campos, SP) (available at: <http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2019/11.21.11.03/doc/publicacao.pdf>)
- Shelia V, Hansen J, Sharda V, Porter C, Aggarwal P, Wilkerson C J and Hoogenboom G 2019 A multi-scale and multi-model gridded framework for forecasting crop production, risk analysis, and climate change impact studies *Environ. Model. Softw.* **115** 144–54
- Souza C M et al 2020 Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and Earth Engine *Remote Sens.* **12** 2735
- Sun S, Bleck R, Benjamin S G, Green B W and Grell G A 2018a Subseasonal forecasting with an Icosahedral, vertically quasi-lagrangian coupled model. Part I: model overview and evaluation of systematic errors *Mon. Weather Rev.* **146** 1601–17
- Sun S, Green B W, Bleck R and Benjamin S G 2018b Subseasonal forecasting with an Icosahedral, vertically quasi-Lagrangian coupled model. Part II: probabilistic and deterministic forecast skill *Mon. Weather Rev.* **146** 1619–39
- Turco M, Jerez S, Doblaz-Reyes F J, Aghakouchak A, Llasat M C and Provenzale A 2018 Skilful forecasting of global fire activity using seasonal climate predictions *Nat. Commun.* **9** 2718
- 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** GB3028
- Vitart F et al 2017 The subseasonal to seasonal (S2S) prediction project database *Bull. Am. Meteorol. Soc.* **98** 163–73
- Wilks D S 2011 *Statistical Methods in the Atmospheric* (New York: Academic)
- WMO 2020 *WMO Guidance on Operational Practices for Objective Seasonal Forecasting* (Geneva: World Meteorological Organization)
- Wood A W, Leung L R, Sridhar V and Lettenmaier D P 2004 Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs *Clim. Change* **62** 189–216
- Zhao W, Zhao X, Zhou T, Wu D, Tang B and Wei H 2017 Climatic factors driving vegetation declines in the 2005 and 2010 Amazon droughts *PLoS One* **12** e0175379
- Zhou X, Zhu Y, Hou D and Kleist D 2016 A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP global ensemble forecast system *Weather Forecast.* **31** 2057–74
- Zhou X, Zhu Y, Hou D, Luo Y, Peng J and Wobus R 2017 Performance of the new NCEP global ensemble forecast system in a parallel experiment *Weather Forecast.* **32** 1989–2004
- Zhu H, Wheeler M C, Sobel A H and Hudson D 2014 Seamless precipitation prediction skill in the tropics and extratropics from a global model *Mon. Weather Rev.* **142** 1556–69