Past and Future of Wildfires in Northern Hemisphere's Boreal Forests

Victor M. Velasco Hererra[1*], Willie Soon[2], César Pérez-Moreno[1], Graciela Velasco Herrera[3], Raúl Martell-Dubois[4], Laura Rosique-de la Cruz[4] Valery M. Fedorov[5], Sergio Cerdeira-Estrada[4], Eric Bongelli[6], Emmanuel Zuniga[7]

[1] *Instituto de Geofísica, Universidad Nacional Autónoma de México, Circuito Exterior, C.U., Coyoacán, CDMX, 04510, México*

[2] *Center for Environmental Research and Earth Sciences (CERES), Salem, Massachusetts, 01970, USA and Institute of Earth Physics and Space Science (ELKH EPSS), Sopron, Hungary*

[3] *Instituto de Ciencias Aplicadas y Tecnología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Coyoacán, 04510, México D.F., México*

[4] *Comisión Nacional para el Conocimiento y uso de la Biodiversidad, Insurgentes Sur 4903, Parques del Pedregal, Alcaldía de Tlalpan, Ciudad de México. C.P. 14010, México*

[5] *Faculty of Geography, Moscow State University, Leninskie Gory St. 1, Moscow 119991, Russia*

[6] *Faculty of Science and Environmental Studies, Department of Geography and Environment, Lakehead University, 955 Oliver Road, Thunder Bay, Ontario P7B 5E1, Canada*

[7] *Instituto de Geografía, Universidad Nacional Autónoma de México, Circuito Exterior, C.U., Coyoacán, CDMX, 04510, México*

Abstract

The boreal forests of the Northern Hemisphere (i.e., covering the USA, Canada and Russia) are the grandest carbon sinks of the world. A significant increase in wildfires could **cause disequilibrium in** the Northern boreal forest's capacity as a carbon sink and cause significant impacts on wildlife and people worldwide. That is why the ability to forecast wildfires is essential in order to minimize all risks and vulnerabilities. We present a

Preprint submitted to Forest Ecology and Management *October 20, 2021*

novel methodology utilizing the Bayesian Machine Learning models to identify climatic variations that induce high and low wildfire activity cycles and forecast long-term occurrences of wildfires. The data analyzed are observed records of wildfires, climate change and climate teleconnections, atmospheric, oceanographic, and environmental factors, starting from the first half of the 20th century. Our Bayesian machine learning models show that a new phase of high wildfire activity in the USA, Canada and Russia began in 2020. While USA has a detectable, oscillation of 40 ± 5 years; Russia and Canada have oscillatory patterns of 30 ± 5 and 60 ± 5 years, respectively. Also, our Machine Learning model forecasts peak wildfire activity at around 2022 ± 3, 2035 ± 3, and 2045 ± 5 **years** for USA, Russia, and Canada, respectively. The new high wildfire activity phase will persist in Russia, USA, and Canada, until 2045, 2030, and 2055, respectively.

Keywords: Wildfires, Environmental Remote Sensing, Machine Learning

1. Introduction

Wildfire is a complex, multi-variable-controlled, emerging phenomenon with the known natural history recorded extending as far back as 450 million years ago (Scott, 2000; Bowman et al., 2009; Rimmer et al., 2015; Doerr and Santin, 2016; Pausas and Keeley, 2019; Zhang et al., 2020). With the arrival of human beings and associated agricultural practices and another large mammals, additional risk factors like land-use changes, landscape modifications (e.g., through animal grazing), and ecological encroachments are all coming into play both active and passive roles in wildfire occurrences and intensities (Marriner et al., 2019; Nanavati et al., 2019; Restaino et al., 2019; Rosan et al., 2019; Schreuder et al., 2019; Williams et al., 2019; Zubkova et al., 2019; Gaboriau et al., 2020).

Both solar magnetic activity and orbital forcing controls on fire events have also been known to be a key factor (Hallett et al., 2003; Daniau et al., 2019; Hamilton et al., 2019; Kappenberg et al., 2019; Glover et al., 2020; Han et al., 2020). In addition, it is also relatively well accepted that abrupt external impact events from asteroids and meteors can be a significant trigger for extensive wildfire and biomass burning (Kennett et al., 2008; Wolbach et al., 2018; Melott and Thomas, 2019; Moore et al., 2019; Pino et al., 2019).

The dynamics and variability of forest fires are related to global and regional climate changes, variations in the atmosphere-ocean circulation and transport modes, and external modulating factors. In addition, there are geographical and ecological settings to consider. There is ongoing development and progress in the seasonal and annual prediction of global forest fire activity (Turco et al., 2018; Shen et al., 2019). However, currently, these forecasts have not enabled the minimization of the ecological deterioration, human and economic losses in the Brazilian and American wildfires in 2019 and 2020, respectively. This is why the prediction of wildfires several years or even a decade ahead is necessary for the security of the Northern Hemisphere's society and a significant scientific challenge. In order to prepare for the danger of wildfires each year, we need to plan and modernizing all environmental contingency programs and early warning systems that will crucially depend on high-quality long-term predictions.

The temperature has been considered one of the main factors in increasing wildfires (Williams et al., 2019; Fletcher et al., 2019). Nevertheless, recently it has been confirmed that the decrease in precipitation may be associated with the increase in forest fires (Williams et al., 2019). It is complicated to imagine that only one climatic variable can explain the wildfire variability; the dynamics and evolution of wildfires in each country within the Northern Hemisphere could involve factors of the atmosphere-ocean circulation and transport modes as external factors such as Total Solar Irradiance (TSI). All these co-factors intervene under different time scales, making the detection and quantification of their roles a significant challenge in the first step of data analyses.

We have analyzed a set of observed records of wildfire, climatic and environmental, and external solar activity parameters to study the nature of the underlying correlations among those variables that can shed light on the fuel-load-hydroclimatic-wildfire mechanisms covering the broad areas of Canada, the USA and Russia. Therefore, to improve our understanding of the complex factors that induce the variability of these wildfires, we have developed a novel algorithm through the powerful techniques drawn from "Machine Learning" as a new tool (Buduma and Locascio, 2017; Ham et al., 2019; Sejnowski, 2020), to understand the complex relationship between the land-atmosphere-ocean system and wildfires in the Northern Hemisphere, which is ultimately essential to permit the prediction of long-term variability of wildfires.

This study aims to identify climatic variations and ecological conditions that induce cycles (high and low activity) of forest fires in the Northern Hemisphere and develop long-term predictions of forest fires.

2. Data and Methods

A wildfire is a fire that spreads uncontrollably and spreads through a forest, rural or urban wilderness vegetation-affecting flora and fauna and wildlife and people-destroying property and deteriorating the environment.

2.1. Satellite Wildfire Data for the Northern Hemisphere

Due to the frequency and magnitude of forest fires in various regions of the world, the use of satellite images has contributed to the detection of hotspots, reducing the response time to the emergency while allowing the analysis of spatio-temporal dynamics of forest fires as a tool to establish the primary factors and elements associated with their occurrence. MODIS (Moderate Resolution Imaging Spectroradiometer) products are currently the most useful and important source of information for hotspot detection because of the advantages shown by this satellite product (Giglio et al., 2016, 2018; Fornacca et al., 2017). They are even used to forecast wildfire activity (Spessa et al., 2015; Ferreira et al., 2020).

The American, Canadian, and Russian land-surface hotspots were obtained from the MODIS Global Monthly Fire Location Product, MCD14ML (collection 6)[1]. This dataset from MODIS Fire SCF at the University of Maryland was selected because of the confidence it provides in the detection of hotspots (Giglio et al., 2016, 2018), since the algorithms and confidence tests used to establish brightness temperature thresholds with the middle and thermal infrared channels and the spatial resolution used (1km) for detection, allowing users to eliminate erroneous pixels (Fornacca et al., 2017). In addition, this dataset provides information and monitoring day/night every minute. The dataset includes descriptive information for each point, such as geographical location, detection date, brightness temperature, radiative energy of the fire, type of inferred heat point (i.e., apparent biomass fire, active volcano, other static ground sources, offshore, and others) and level of trust/confidence.

We used for the analysis all hotspot points detected from 01/11/2000 to 30/06/2020 because the seasonality of wildfires in the United States, Canada and Russia occurs throughout the year. The data are downloaded in shapefile format for processing in a geographic information system (GIS). We eliminated those hotspots defined

[1] https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/mcd14ml

as static: volcanoes, industries, oil wells, and anthropogenic activity in the USA, Canada, and Russia. Also, those detected with a confidence level of less than 75% we eliminated. Therefore, we strictly study satellite data related to wildfires.

2.2. Historical Wildfire Data for the Northern Hemisphere

We will analyze the historical data of the following countries in the Northern Hemisphere, chosen for being the most important carbon sinks in the world: a) American wildfires (1926-2020)[2], b) Canadian wildfires (1930-2020)[3],[4] and c) Russian wildfires (1950-2020)[5],[6].

We would like to highlight that all the historical records of forest fires are incomplete, which have made their analysis uncertain. This is why we carry out a Bayesian analysis that allows us to find a model that describes the variations of forest fires in the USA, Canada and Russia probabilistically in order to account for the incompleteness of the available historical records.

2.3. Climate Teleconnections

The next set of annual time series we used is from the National Oceanic and Atmospheric Administration[7]: 1) Accumulated Cyclone Energy (ACE),

[2] https://www.nifc.gov/
[3] https://cwfis.cfs.nrcan.gc.ca/ha/nfdb
[4] https://www.ccfm.org/
[5] http://rosleshoz.gov.ru/
[6] https://doi.org/10.1007/978-94-015-8737-2-8
[7] https://www.esrl.noaa.gov

2) Arctic Oscillation (AO), 3) Atlantic Multidecadal Oscillation (AMO), 4) North Atlantic Oscillation (NAO), 5) Pacific Decadal Oscillation (PDO), 6) Palmer Drought Severity Index (PDSI). Also, we used 7) El Niño/Southern Oscillation (ENSO)[8], 8) World temperature and precipitation data[9], 9) Total Solar Irradiance (TSI)[10].

In order to weigh and inter-compare the variables analyzed in the study of Northern Hemisphere wildfires, we adopted the standardized annual data, i.e., with zero average value and unit standard deviation.

2.4. Multiple-time-series Cross Wavelet Spectrum

We have used MATLAB 2019b, the Wavelet Toolbox, the cross wavelet and wavelet coherence toolboxes for MATLAB by Grinsted, Moore and Jevrejeva (Grinsted et al., 2004), the Torrence & Compo Wavelet Analysis Software (Torrence and Compo, 1998) and our new Multiple Cross Wavelet alogrithms (Velasco Herrera et al., 2017; Soon et al., 2019). The main goal of our data analyses is to find the possible climatic patterns and factors responsible for the underlying cycles in Northern Hemisphere wildfires. There are different methods to find patterns in time series. We use the wavelet analysis because this method allows identification of the intricate patterns of the phenomenon (such as wildfires) and the patterns of interaction with associated co-factors (Soon et al., 2019).

[8] https://www.pnas.org/content/116/45/22512
[9] https://climateknowledgeportal.worldbank.org/
[10] https://doi.org/10.7910/DVN/SURA99

Wavelet transform (see e.g., Grinsted et al., 2004; Torrence and Compo, 1998; Velasco Herrera et al., 2017; Soon et al., 2019) can be considered as an intelligent system and is applied here to find patterns (periodicities), its evolution in the time, as well as to make predictions. Furthermore, it can also be used as an optimal filter.

We applied our new Multiple-time-series Cross Wavelet spectrum ($\mathbf{\Omega}^{\otimes}$, Velasco Herrera et al., 2017) in order to identify climatic patterns and ecological conditions that induce high and low cycles in wildfire activity. Our Multiple-time-series Cross Wavelet spectrum (MCW) is based on the generalized Einstein's cross function ($\mathfrak{M}$) (Velasco Herrera et al., 2017). The relationship between $\mathfrak{M}$ and $\mathbf{\Omega}^{\otimes}$ are the following:

$$\mathbf{\Omega}^{\otimes} = \mathbf{W}[\mathfrak{M}] \tag{1}$$

$$\mathfrak{M} = \mathbf{W}^{-1}[\mathbf{\Omega}^{\otimes}] \tag{2}$$

where $\mathbf{W}$ and $\mathbf{W}^{-1}$ is the wavelet transform and inverse wavelet transform, respectively. The $\mathbf{\Omega}^{\otimes}$ spectrum is defined as the product (**Track**) of the diagonal elements in spectral wavelet hipermatrix ($\mathbf{\Omega_{total}}$) and is given by the formula (Velasco Herrera et al., 2017):

$$\mathbf{\Omega}^{\otimes} = \mathbf{Track}\left(\mathbf{\Omega_{total}}\right) = \prod_{i=1}^{i=n} \mathbf{\Omega_{total}}_{ii} = \langle \mathbf{W}_{11} \otimes \mathbf{W}_{22} \otimes \ldots \otimes \mathbf{W}_{nn} \rangle_{[t,s]} \quad (3)$$

where.

$$\mathbf{\Omega_{total}} = \begin{pmatrix} \langle \mathbf{W}_{11} \rangle_{[t,s]} & 1 & \cdots & 1 \\ 1 & \langle \mathbf{W}_{22} \rangle_{[t,s]} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \langle \mathbf{W}_{nn} \rangle_{[t,s]} \end{pmatrix}$$

and $\langle \circ \rangle_{[t,s]}$ indicates for the wavelet spectrum smoothing in both time (t) and scale (s).

So that the $\mathbf{\Omega}^{\otimes}$ spectrum (Equation 3) is different from zero is necessary that all time series have at least the same frequency. This implies a synchronization of the land-atmosphere-ocean system with the Northern Hemisphere wildfires at the same frequencies. In this way, the climatic patterns that induce high and low cycles in wildfire activity will be found.

MCW has an intelligent algorithm to simultaneously analyze "N" variables (N $\geq$ 2) and find the complex or linear relationships that exist between all the variables. We use the Morlet wavelet basis in the MCW because it has among the highest precision in resolving the periodicities that all "N" time series have in common and because it is a complex function that allows us to obtain the information on phase as well, that is represented by arrows in the figures of the main text.

The inputs in the MCW are the “N” time series and MCW has 4 outputs as shown in Figs. 3, 5, and 7 below: *i*) The global frequency spectrum (or time-averaged), which shows the periodicities (patterns) existing in all the “N” variables (left panel). *ii*) The local spectrum, that shows the evolution over time of these periodicities as well as their phase (center panel). *iii*) Global phase, shows the average phase of the “N” variables (right panel) and *iv*) Multi-cross function, amplitude and phase, of the dominant pattern (bottom panel).

2.5. Machine Learning Algorithms for Probabilistic Forecasting of the Northern Hemisphere Wildfire Activity

Historical wildfire data has uncertainty, so it is important to select a Machine Learning (ML) model that is able to adequately approximate the wildfire dataset with a high level of confidence. There are several ML algorithms, and we selected Bayesian inference machine learning for our purpose. Also, we will use the Bayesian inference ML model obtained from each of historical wildfires records as input to the Least Squares-Support Vector Machines (LS-SVM, see Suykens et al., 2002) algorithms to obtain probabilistic models of forecasting Northern Hemisphere wildfire variability beginning in the year 2020 AD. Also, we note that any ML model is limited by an uncertainty principle.

Non-linear Autoregressive EXogenous (NARX) model

In order to create forecasting models of wildfires activity, we use the Nonlinear Autoregressive EXogenous model ($\widehat{\mathbf{y}}$) that is defined as (Vapnik, 1998; Suykens et al., 2002):

$$\widehat{\mathbf{y}}_{[t+1]} = \boldsymbol{f}(y_{[t,p]}, u_{[t,q]}) \tag{4}$$

where $\boldsymbol{f}$ is a a non-linear transfer function that depends on the input (y) and output (u) data, and p and q represent the number of lags of the input and output values, respectively. So, $\widehat{\mathbf{y}}$ is the estimated wildfire time series at time "$t+1$".

Bayesian inference for LS-SVM regression

To create probabilistic models of the wildfires activity, we use a Bayesian inference model obtained from the wildfire records for each country analyzed (USA, Canada, and Russia). Bayes's theorem is the basis of these models and can be expressed as follows:

$$Posterior(\boldsymbol{f}|D) = \frac{Likelihood(D|\boldsymbol{f})}{Evidencep(D)} Priorp(\boldsymbol{f}) \tag{5}$$

where D is training data, in our case is the wildfire records and $\boldsymbol{f}$ is the Least-Squares Support-Vector Machines (LS-SVM) regression model:

$$\boldsymbol{f} = \sum_{t=1}^{n} \boldsymbol{\alpha}^{\mathbf{t}} K(u, u_t) + \boldsymbol{\beta} \tag{6}$$

where $\boldsymbol{u}_t$ is the value of the Bayesian inference model of the wildfires at time "t" (discrete time index from $t = 1, \cdots, n$), K is the kernel, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are hyperparameters. The output is the estimated value of $\widehat{\mathbf{y}}$.

Bayes's theorem is used to deduce the optimal hyperparameters of the LS-SVM model (see Suykens et al., 2002, for technical questions about the method).

2.5.1. Algorithms for the estimation of wildfire cycles

In order to forecast the next high wildfire season in the USA, Canada, and Russia, we apply the following iterative steps:

I. Use multiple cross-wavelet transform (Equation 3) to find the periodicities on climate teleconnections and wildfires record for each country, i.e., USA, Canada, and Russia. The results are shown in Figures 3, 5, and 7.

II. Use Equation (5) to obtain a Bayesian inference model from the time series of wildfires records (USA, Canada, and Russia) and shown as blue lines in Figures 4, 6 and 8.

III. Selection of the model lags "p" and "q" for each Bayesian inference model that has been analysed (USA, Canada, and Russia).

IV. Use the K-fold cross-validation for the training, validation, testing and deduction of the parameters of the NARX model (Equation 6).

V. Set aside 1/K of data. Train the model with the remaining (K-1)/K data. Measure the accuracy obtained on the 1/K data that we had set aside. K independent training is therefore acquired. The final accuracy will be the average of the previous K accuracies. Note that we are "hiding" a 1/K part of the training set during each iteration. This is applied at the time of training. After these K iterations, we obtain K accuracies that should be "similar" to each other; this would be an indicator that the model is working well or not. In this work, we used K=10, but, is possible to vary K between 5 and 10.

VI. Use Bayes's theorem to deduce the optimal hyperparameters (α and β) of the LS-SVM model (Equation 6).

VII. Estimation of the following high and low wildfire activity cycles using Eq. (6).

VIII. Computation of a cost function.

IX. Test of the accuracy of the estimated wildfire activity cycles.

X. Test of the cost function: if this function was small enough, we stopped and went to the next step (XI). Otherwise, we change one of the parameters and repeat from step (III) onwards.

XI. Use the wavelet transform to help determine if the periodicities of the estimated wildfire cycles have the same periodicities obtained in (I). If yes, then done and accept the estimate. Otherwise, repeat the estimate from step (III).

We have used and modified the LS-SVM algorithms and toolbox by Suykens et al. (2002) to forecast the next high wildfire season in the USA, Canada, and Russia. The LS-SVMlab toolbox contains Matlab/C implementations for a number of LS-SVM algorithms by J.A.K. Suykens The LS-SVMlab software is made available for non-commercial research in `https://www.esat.kuleuven.be/sista/lssvmlab/`.

3. Results

3.1. Spatial Analyses of the Northern Hemisphere Wildfires

Land cover information is integrated and processed with hotspot data in a geographic information system (GIS) to establish fire frequency and vulnerability percentages to determine the vegetation most vulnerable to forest fires in the regions of Canada, the United States, and Russia. The information generated is associated with the orography of the terrain, hence allowing us to obtain dominant altitudinal values. The results are shown in Figures 1 and 2.

The hotspot data used (red dots in Figures 1a and 2a) are from the MODIS Collection 6 series: Temporal coverage is from 2000 to 2020 with

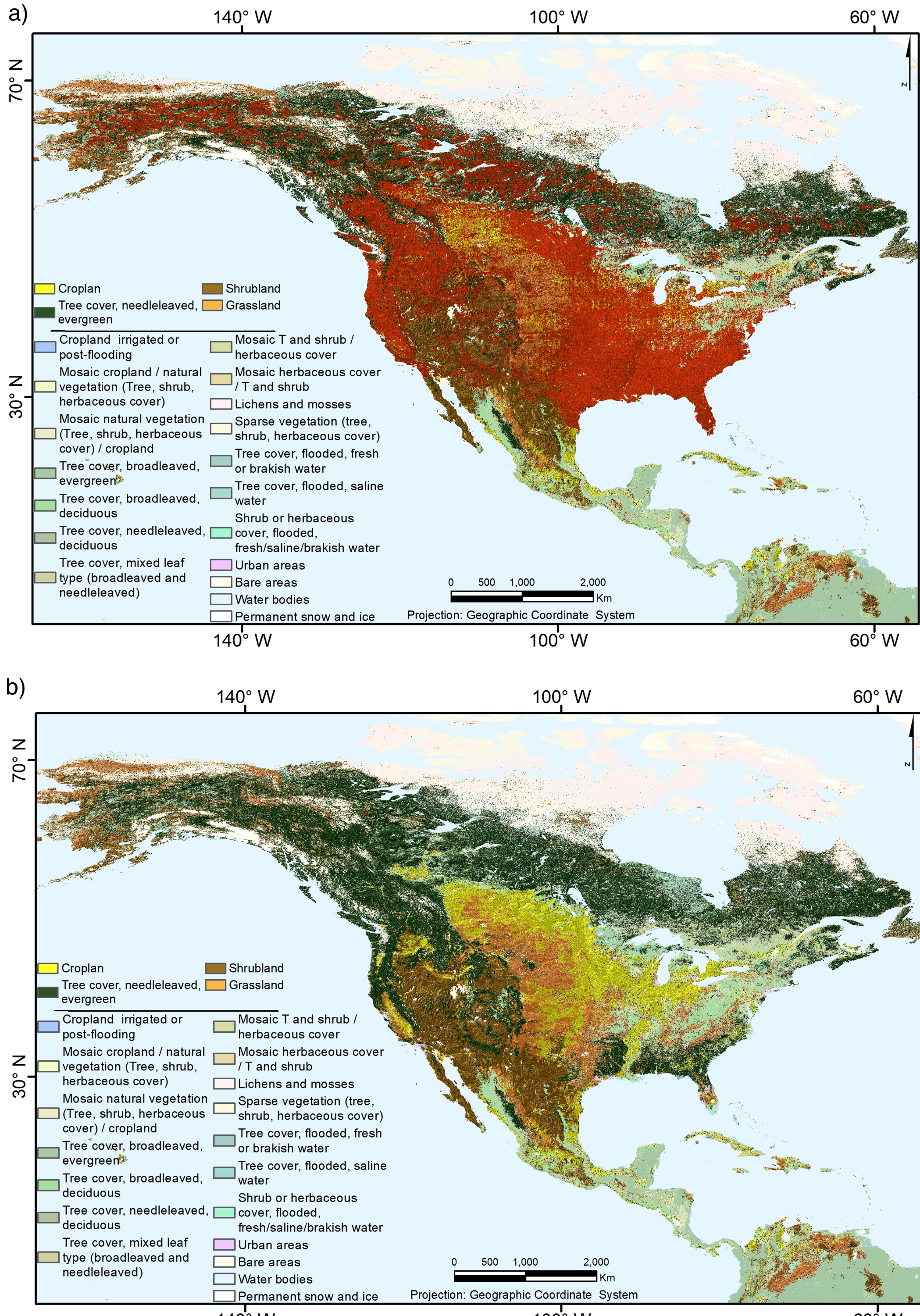

Figure 1: a) The North American land cover map superposed with the spatial distribution of satellite wildfire hotspots data (red points) in USA and Canada from 2000 to 2019 is shown. b) North American vegetation cover most affected by forest fires in the USA and Canada: Cropland (yellow), 2) Shrubland (brown), 3) Tree cover, needle-leaved, evergreen (green) and 4) Grassland (orange)

a confidence level greater than 75%, eliminating information from active volcanoes and other static sources on land and offshore.

The Land Cover (LC) information is obtained from the Climate Change Initiative (CCI) project of the European Space Agency (ESA). Global maps represent these geospatial data in raster format with a spatial resolution of 300 m, classified into 22 types of coverage and corresponding to the time interval from 1992 to 2018. **The CCI-LC (ESA) data set is represented by global LC images with a spatial resolution of 300 m and an annual resolution from 1992-2020. The products provide 38 types of LC classified based on the typology established by the Food and Agriculture Organization of the United Nations (FAO)[11]. This Land Cover Classification System (LCCS) based on numerical codes was converted to LC information for the 2000-2020 periods and LC types for the United States, Canada and Russia. From the location of the wildfires, we got 22 types of LC (see Table B1).**

To determine altitudinal levels, topographic data from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) are used, with a resolution of 7.5 arc seconds (225 meters). To quantify the impact of forest fires on the different land covers in the United States, Canada and Russia, the data between 2001 and 2018 are compared **(see for example Liu et al., 2019).**

[11]https://www.esa-landcover-cci.org/

The results show that in the Northern Hemisphere, the most significant impact due to the increase in forest fires is related to the vegetation category of tree cover, needle-leaved, evergreen land covers. Table 1 shows the main vegetation covers affected in the USA, Canada and Russia by the percentage of the total number of forest fires in 2001 and 2018.

It should be noted that under the vegetation category of tree cover needle-leaved and evergreen in the USA, the percentage of the number of forest fires affected has increased from 37.2% in 2001 to 46.1% in 2018.

In the 17 years interval, there is an increase of 8.9%. This contrasts sharply with a slight decrease or small increases in the fire-affected vegetation types of Cropland, Shrubland and Grassland for the same period (see Table 1 for more details). For Canada, under the vegetation category of tree cover needle-leaved, and evergreen, the number affected by forest fires has increased from 31.3% in 2001 to 77.8% in 2018. Therefore, an increase of 46.5% had been recorded for the same period of 17 years. At the same time, this substantial increase in wildfires affecting this vegetation cover can be contrasted with the decreases under the other three vegetation categories of tree cover-mixed-leaf type (broad-leaved and needle-leaved), Cropland and Shrubland (see Table 1). Finally, in Russia, the fire affected the vegetation type of tree cover, needle-leaved, evergreen has increased from 36.1% in 2001 to 55.7% in 2018.

This registers an increase of about 19.6%. While the fire-affected vegetation types in Russia under cropland, grassland and tree cover, broad-leaved, deciduous covers decreased or slightly increased over the same 2001-2018 interval (see Table 1).

Table 1: Percentage of forest fires in the main land covers of the USA, Canada and Russia

Canadian vegetation cover most affected by wildfires	2001	2018
Cropland rainfed	31.5%	4.3%
Tree cover, needle-leaved, evergreen	31.3%	77.8%
Shrubland	9.4%	3.6%
Grassland	3.1%	1.2%
American vegetation cover most affected by wildfires	**2001**	**2018**
Tree cover, needle-leaved, evergreen	37.2%	46.1%
Shrubland	14.7%	15.3%
Cropland rainfed	14.4%	9.6%
Grassland	9.3%	9.3%
Russian vegetation cover most affected by wildfires	**2001**	**2018**
Tree cover, needle-leaved, evergreen	36.1%	55.7%
Cropland rainfed	28.4%	10.1%
Grassland	8.2%	6.6%
Tree cover, broad-leaved, deciduous	6.1%	9.3%

The satellite recorded values of the brightness temperatures (hotspots) will depend on the type of vegetation and trees, the humidity and water conditions of the vegetation-mass fuel and the number of burned trees during fires. The type of tree, in turn, also depends on climatic and geographical conditions. GIS information in Figures 1 and 2 shows that when analyzing the brightness temperatures of fires from 2000 to 2020, the difference of wildfires in the plains and the mountains is clear. The differences in the tundra and desert areas are also clearly distinguishable. Fig. 1b and 2b

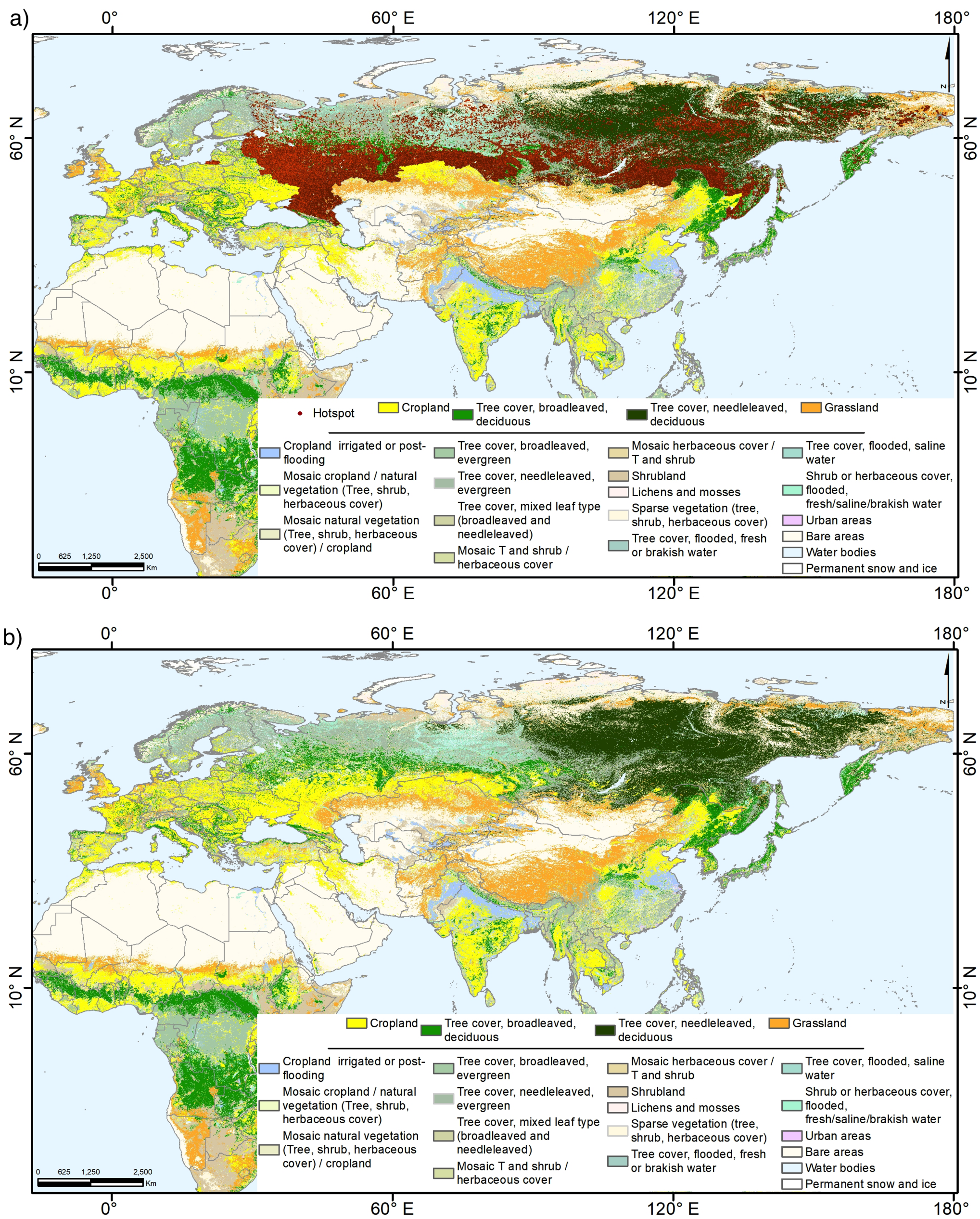

Figure 2: The Northern Eurasia land cover map. a) The Northern Eurasia land cover map superposed with the spatial distribution of satellite wildfire hotspots data (red points) in Russia from 2000 to 2019 is shown. b) North Hemisphere vegetation cover most affected by forest fires in the Russia: Cropland (yellow), 2) Tree cover, broad-leaved, deciduous (light green), 3) Tree cover, needle-leaved, evergreen (strong green and 4) Grassland (orange)

show the clusterings' results, and we can classify four different land covers where more than 75% of forest fires occur in the USA, Canada and Russia.

Every year wildfires, no matter the multivariate causative agents, indeed severely affect the human society and the environment in the Northern Hemisphere alike. Various public policies, ranging from active management preparedness to emergency responses, have been leveled to allow humanity and natural ecological environment to cope with the danger of fire. Therefore, any promise for a long-term prediction of wildfire occurrences is not only an urgent but also a powerful capability that can help to minimize the risks and vulnerabilities of Northern Hemisphere's society from wildfires. In addition, the results of GIS illustrated here can help select specific forests/vegetations for monitoring climatic conditions, particularly rainfall and drought and soil moistures. Such kind of intelligent information gathering and processing will allow particular measures to minimize economic, human and ecological losses before a fire begins in any vulnerable areas.

3.2. Tools for Understanding and Predicting Frequency of Wildfires in USA, Canada, and Russia

The GIS clustering analysis shows the spatial variation of the Northern Hemisphere wildfires. In the spatial sense, each cluster obtained recognizes very well defined regions and delimited areas. This may allow one to optimally plan to minimize the risks and vulnerabilities of Northern Hemisphere society from wildfires by setting local and regional management priorities.

However, this complexity does not prevent or paralyze the narrower study of wildfire time-series statistics in each country's analyzed (i.e., Canada, USA, and Russia). Using Machine Learning, we propose a new methodology to make long-term, several decades-long, forecasts for the wildfires in the USA, Canada and Russia.

- American wildfires

To begin studying the complex relationship between the forest fires of the Northern Hemisphere and the land-atmosphere-ocean system, we will analyze the American wildfires. In this first case, twelve variables are assessed (N=12) in the MCW, and these time series are shown in the top panel of the Figure 3: 1) the number/frequency of American wildfires, 2) Burned Area, 3) PDSI, 4) surface temperature, 5) precipitation, 6) snow cover, 7) AMO, 8) PDO, 9) NAO, 10) ACE, 11) ENSO, and 12) TSI.

The global time-averaged MCW shows two significant patterns (periodicities) at decadal-10 years and multi-decades 40±5 years, with more than 95% confidence level (dotted red line, left panel) in American wildfires due to the combined modulation of the land-atmosphere-ocean system and the total solar irradiance (TSI). The decadal periodicity and its relative persistence are most likely related to the solar activity cycle and its teleconnections in climatic signals. The spectral power of this periodicity is present in the entire time interval (1926-2019), being more intense from1935 to 1955 and between 1975 to 1995. We further note that the maximum values in the decadal spectral power are timed around the maximum of the multi-cross function of the multidecadal scale around 40 years (blue curve in the bottom panel). The local decadal and 40-year multidecadal phases do not show a well-defined orientation (that is, the arrows point in different directions), so the relationship between wildfires and the atmosphere-ocean system is complex. This fact can be reconfirmed through the behaviour of the global phase time-averaged result plotted in Figure 3 (black line, right panel).

Despite the complexity of this system, the 40-year multi-cross function is theoretically in phase and in time equivalent of all climatic indices and American wildfires. This fact will allow the use of this function to extrapolate to future scenarios, make theoretical forecasts on the tendency of American wildfires, and then compare it with the predictions obtained with the Machine Learning method discussed below.

We note that climatic oscillations with multi-decadal periodicities have been reported in many previous works (e.g. Soon, 2009; Soon et al., 2015; Le Mouël et al., 2019). The variations in the NAO, AMO and PDO have a strong impact on climate variability in sea-surface temperature, air temperatures, rainfall, precipitation, stream flow, and surface temperature anomalies of North America (e.g. Kitzberger et al., 2007; McCabe et al., 2008; McCabe-Glynn et al., 2013; Soon et al., 2015; Le Mouël et al., 2019).

In particular, it is of great interest to know the influence of ENSO on annual and multi-year variations in wildfires statistics. We first note that the imprints of the El Niño phenomenon do not always show up in the increase

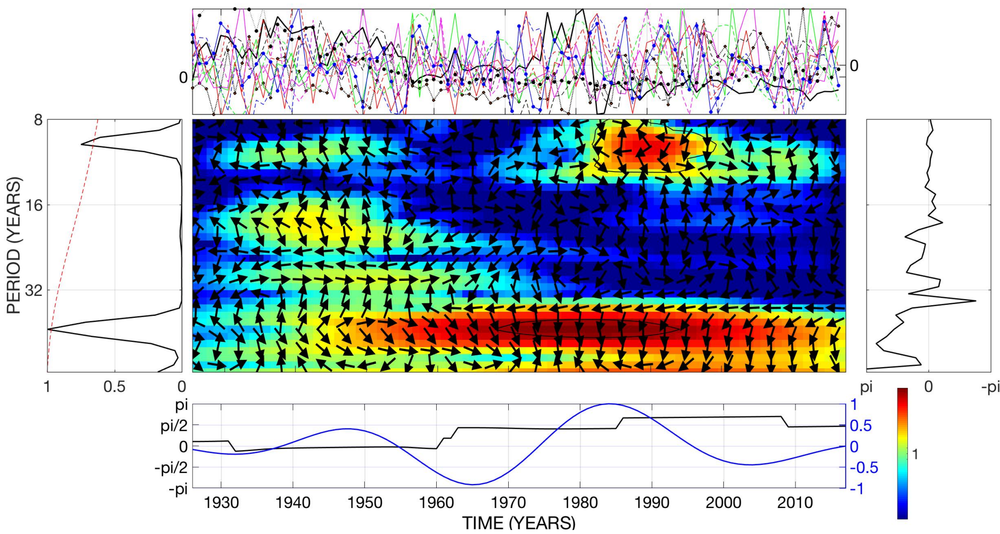

Figure 3: Time-frequency multi-cross wavelet from 1926 to 2019 between number of American wildfires, burned area, surface temperature, precipitation, snow cover, atmospheric-oceanic circulation and energy indices PDSI, AMO, PDO, NAO, ACE, ENSO, and the external solar forcing factor TSI. In the central panel, the calculated local wavelet power spectral density (LWPSD) in arbitrary units is shown adopting the red-green-blue colour scales. The black arrows indicate the relative phase of the synchronization. The orientations from left to right (→) and from right to left (←) indicate that there is a linear, in-phase or antiphase, synchronization at a certain frequency between all time series. Any other orientation means that there is a complex, non-linear synchronization. The bottom panel shows the multi-decadal cross function at the significant timescale of 44-years (blue line) and the instantaneous phase relative for the same multi-decadal oscillation (black line). The global time-averaged wavelet period is shown in the left-hand panel with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The panel on the right shows the global time-averaged phase values.

in the number of forest fires nor the increment in the burned area. The historical data of the American wildfires also show annual and multi-annual variations and the decadal and multidecadal signatures. MCW analysis does not show these periodicities suggests that these annual and multi-annual patterns have only local effects on the forested areas in America and that the seasonal atmosphere-ocean climate conditions may be more dominating.

From the point of view of signal theory, the absence of annual and multi-annual variations means that they are considered as noise. Therefore, to predict forest fires, we should not focus on predicting these annual and inter-annual variations. In sharp contrast, the decadal and multidecadal periodicities result from the more persistent and coherent interactions of the coupled solar-land-atmosphere-ocean system. Because that is essentially a highly variable and stochastic process, it is impossible to say precisely the number of forest fires for the following years.

The objective of using Bayesian Machine Learning models is to give an interval in which the number of wildfires can vary, with a high confidence level ($> 95\%$). Also, we used the average value and the standard deviation of the historical data of American wildfires (which we called "objective data") to quantify and define when there are high and low cycles of the frequency of American wildfires.

To support our choice, we show a comparison (Fig. 4) between the objective data (historical data of the American wildfires in black line) and the Bayesian Machine Learning model (blue line) from 1926 to 2020. This model represents the high and low-frequency fluctuations of American wildfires. It is observed that the objective data is indeed well distributed around the Bayesian Machine Learning model.

With the support of the mean value (horizontal solid black line) and the standard deviations σ^+ and σ^- (a standard deviation above the mean value and a standard deviation below the mean value black dotted lines, respectively), we note that because the maximum values of the objective data (black line) are above the standard deviation σ^+ from about 1930s-1950s and 1970s, these events can be classified as severe wildfire phase. The first minimum of this objective data occurs between 1950 and 1970, and it is around the average value so that this period can be classified as a moderate wildfire phase. While the second minimum is between 1985 and 2005, it is below and around the standard deviation σ^-, so this period can be classified as a low wildfire interval.

We note that the objective data is oscillating around the multidecadal Bayesian Machine Learning model (i.e., trend), which implies that the decadal variations are modulated by the tendency of American wildfires that were co-generated by the weather/climate/ecological conditions.

There are several techniques to make time series predictions (Kubat, 2015). Each of these methods have favorable aspects as well as their weaknesses. Once we have obtained the Bayesian Machine Learning model that show the occurrence of high and low cycles in forest fires, it is now possible to select a Machine Learning algorithm to make a prediction of American wildfires that is based on their decadal and multidecadal co-patterns.

Before forecasting the number of wildfires for the following decades, it is necessary to quantify the ability of the Bayesian model to "predict" a variation in the recent and past wildfires. We use 80% of the Bayesian model (that is, data from 1926 to 2001) as input data to "forecast" the remaining 20% of the Bayesian model (i.e., 2002 to 2019). The Bayesian model of the historical data shows that all the annual historical data oscillate around the Bayesian model; this fact indicates no overtraining or undertraining. Furthermore, the multiple cross wavelet analysis shows that the high and low seasons of forest fires have a multidecadal variation, so the Bayesian model we deduced is not overly complex, which implies that the validation is simple. We do not show the validation figures but instead choose to concentrate on the forecasting result.

Based on the Bayesian model obtained from American wildfires, we have selected the Least Squares Support Vector Machines (LS-SVM) with the Nonlinear Autoregressive Exogenous Model (NARX, see Vapnik, 1998; Suykens et al., 2002, for more details about method) to predict the next few cycles of American wildfires.

We used the Bayesian model from 1926 to 2020 obtained by the objective data to train the LS-SVM. Once those trainings are completed, we obtain the prediction model that would show the probabilistic forecast of the activity of

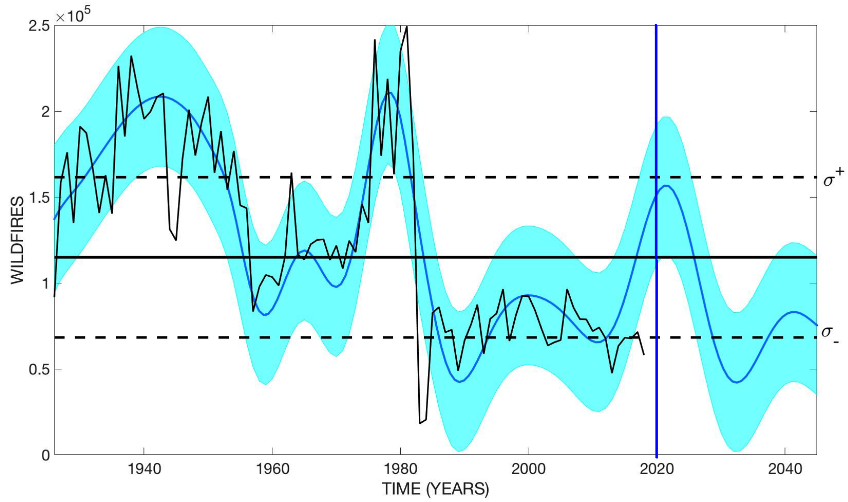

Figure 4: Annual frequency of wildfires in USA from 1926 to 2019 (black line) compared with the Model's Machine Learning Bayesian inference (blue line). The horizontal solid black and dashed black lines are the mean fire frequency and its one standard deviation respectively, for the objective data from 1926-2019 interval. The blue shaded area represents the 95% confidence intervals of the Bayesian ML model.

American forest fires between 2021 and 2030. The validity of the prediction model was assessed with K-fold cross-validation (in this work, we adopt $K = 10$). It was indeed necessary to evaluate how we would optimally combine the models obtained by the Bayesian Machine and the LS-SVM models. For that, it is necessary to look for a correction function in order to calibrate the predictions.

There are different calibration methodologies, and we select a calibration that homogenizes and standardizes all measurements of the models obtained (see Soon et al., 2019, for more details). In addition, this methodology allows us to continue using the average value and standard deviations as a criterion to quantify the next cycle of forest fires. After calibrating the forecasting model, we again used Bayesian Machine Learning to obtain a probabilistic model of American wildfires.

The results obtained from the Bayesian prediction model are shown to the right of the vertical blue line in Fig. 4. The blue shaded area represents the 95% confidence intervals of the Bayesian ML model. The results obtained from the prediction in Fig. 4 show that a new high cycle of forest fires has begun and could last for the next 4 to 7 years. In addition, this new cycle, by being in between the average value and the standard deviation σ^{+}, can be classified as moderate to severe wildfire conditions.

The fire will be probably manifest in all American wild forests, and other landscapes and the American burned areas could be well above those from the last 20 years. Such a future scenario could cause severe ecological, environmental damage with significant human and economic losses. But in the mean time, Fig. 4 predicts that around 2040, there will be a low cycle of forest fires in America comparable to those low fire regimes that occurred between 1980 and 2010.

Once the model explaining intrinsic patterns, that is, multi-decadal oscillation of wildfires, have been obtained, it is now possible to explain the complex evolution of the historical number of forest fires from their interaction with climatic variations, ecological conditions, atmosphere-ocean circulation and transport modes as well as external factors such as solar TSI. The high cycles of American wildfires (1926 to 1955, 1970 to 1990 and likely 2019 to 2030) is because of a prolific decrease during those years, well below its average value of precipitation, snow and ACE. This persistent condition causes a prolonged and severe drought. In addition to a positive phase of the PDO and the ENSO causes a warmer climate and dry air, therefore an increase in air temperature. Also, there is less cloudiness and lower atmospheric humidity that causes greater penetration of the solar radiation to the ground or near-surface. All these multiple co-factors cause a considerable accumulation of dry biomass fuels, and therefore both a combination of natural and human factors cause a large number of forest fires causing an extensive burned area of forests.

Low wildfire cycles (1955-1975, 1990-2018) mainly were likely attributable to an increase in rainfall, snow and ACE well above its average value, as well as a negative phase of the PDO. All such conducive conditions cause a wetter climate. In addition, most of the dry biomass fuels were previously burned. During such periods, forests and ecosystems underwent a recovery and growth of vegetation and trees. Until a new high cycle of wildfires recommences.

Concerning, the annual and multi-annual variations of ENSO and its effects on wildfires, it can now be explained that its effects contribute to the increase in the number of forest fires and to the increase in burned areas, when these variations occur at the maximum of the multi-decadal oscillation. During ENSO occurrence around the minimum phase of this oscillation, its effects are practically neutralized and absent.

- Canadian wildfires

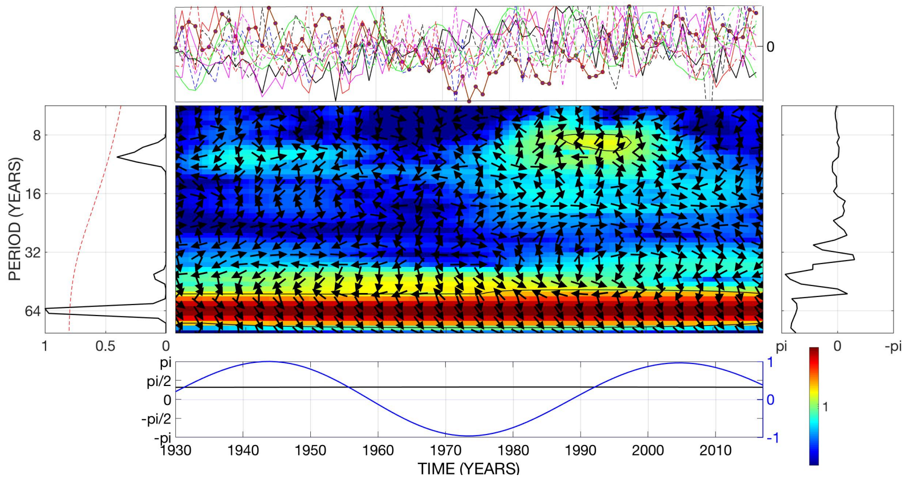

Figure 5: Time-frequency multi-cross wavelet from 1930 to 2019 between number of Canadian wildfires, burned area, surface temperature, rainfall, snow cover, atmospheric-oceanic circulation and energy indices AMO, PDO, NAO, ENSO, ACE, and the external solar forcing factor TSI. The bottom panel shows the multi-decadal cross function at the significant timescale of 60-years (blue line) and the instantaneous phase relative for the same multi-decadal oscillation (black line). All other panels present similar information as described in Fig. 3 but for the Canadian wildfire statistics.

Canada has also compiled an excellent historical record of wildfires (1930-2020), and we used a second MCW to find patterns in their wildfires caused by co-variations in AMO, NAO, PDO, ENSO, ACE, TSI, burned area, rainfall, snow cover and their surface temperatures. For this second case of wildfires, we have N = 11, and these time series were also standardized to be used in the input data in the MCW (top panel in Fig. 5). The global wavelet spectrum shows a decadal and a multidecadal pattern of 60±5 years again.

The first pattern is slightly below the 95% confidence level, but the second multidecadal period is above.

The local spectral power of the decadal pattern (center panel) is around the maximum of the 60-year multi-cross function (blue curve in the bottom panel) and the phase for this decadal modulation does not have a definite tendency (the arrows are in all directions), so the relationship between Canadian wildfires, atmosphere-ocean system and decadal TSI is complex. While the 60-year multidecadal pattern phase has a quasi-perpendicular orientation over the entire time interval, its spectral power/signal is very high.

The number of Canadian wildfires shows more significant inter-annual variability than American wildfires. That could be due to the relatively more extreme climatic conditions/oscillations to which Canadian forests and landscapes were subjected at higher latitudes. In addition, it is again observed that the co-factor El Nino influences the more excellent dispersions of the inter-annual data during the positive phase of the 60-year multidecadal oscillation.

Fluctuations with an average period of 60 years are known in different hydrometeorological processes. This oscillation is reported in the processes of the ocean-atmosphere system and the variability of the surface air temperature. As well as in the dynamics of the sea ice area in the northern hemisphere (Leal-Silva and Velasco Herrera, 2012; Fedorov, 2018).

The 60-year oscillation is most clearly manifested in the North Atlantic (Fedorov, 2018). It has been suggested that the Earth's rotation is one of the modulating sources of different hydrometeorological processes and, in particular, in the 60-year periodicity. Nevertheless, there is a discrepancy if it is the ocean-atmosphere system or it is cosmic in nature (Soon et al., 2011, 2014; Fedorov, 2018) and is plausibly related to the solar barycentric motion

(Cionco and Soon, 2015; Cionco and Pavlov, 2018) that cause the variations in the dynamics of the Earth's rotation. However, there is still no consensus on the genesis of this periodicity.

Additionally, the global wavelet spectrum indicates a relatively weak periodicity of 40 years that is well below the 95% confidence level. This pattern is however manifested within the American wildfire statistics and could represent a latitudinal relationship of the forests of Southern Canada with the forests of the Northern USA. In addition, clear inter-annual fluctuations are also absent in the MCW because they are patterns of each of the Canadian wildfire regimes/zones and not all of these regional wildfires are synchronized with the global circulation indices and TSI co-factor when wildfires occur.

We use again the Bayesian Machine Learning to obtain model (blue solid curve in Fig. 6) that describe the variability of Canadian wildfires between 1930 and 2020 (i.e, objective data), which are represented as a black line to the left of the vertical blue line of Fig. 6. It can be noted again that the historical annual-based data of these fires are well distributed around the Bayesian model. This model represents the multi-decadal frequency fluctuations of Canadian wildfires.

Canadian wildfires show a very low activity season between 1930 and 1965 (despite having very hot summers during the 1930s), below the standard deviation (lower horizontal blue dotted line in Fig. 6) of the entire record. During this period, a very high accumulation of ice has been reported in the Northern Hemisphere. This may have caused a very low cycle phase of the

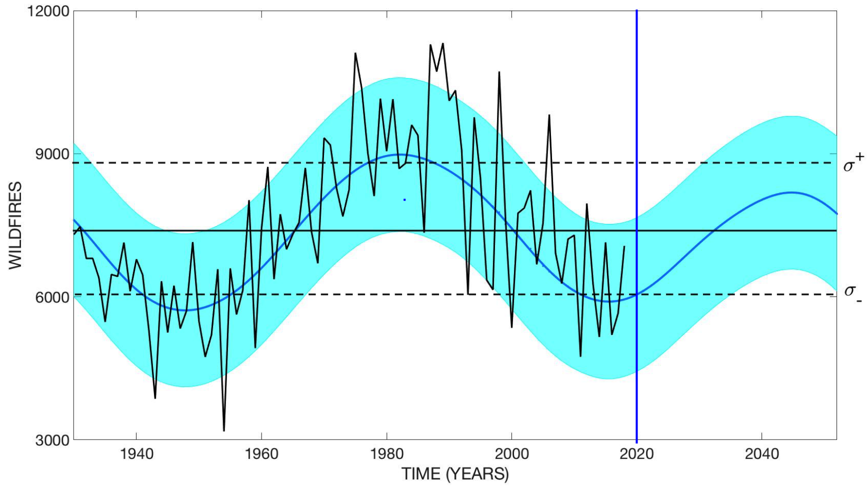

Figure 6: Annual frequency of wildfires in Canada from 1930 to 2019 (black line) compared with the Model's Machine Learning Bayesian inference (blue line). The horizontal solid black and dashed black lines are the mean fire frequency and its one standard deviation interannual statistics, respectively, for the objective data from 1930-2019 interval. The blue shaded area represents the 95% confidence intervals of the Bayesian ML model.

Canadian wildfires. In contrast, from 1970 to 1990, there was a very high season of wildfires with an extended duration above the standard deviation (upper horizontal blue dotted line in Fig. 6). During 2010 and until 2020, there is a very low season of Canadian wildfires since the values are generally below the standard deviation.

For the prediction, we used a new LS-SVM and trained with the Bayesian model obtained from the objective data between 1930 and 2020. After calibrating the forecasting model, we again used Bayesian Machine Learning to obtain a probabilistic model of Canadian wildfires. The prediction of the Canadian wildfire activity was validated again with the K-fold cross-validation ($K = 10$). It can be seen that the next maximum Canadian wildfire obtained by the Bayesian method is timed around 2040, and it is predicted that it will be a severe cycle of Canadian wildfires with activity above the standard deviation of the wildfire statistics.

From 2021-2022 onward, the number of fires will grow every year and most likely, after 2025, the number will be above the historical average value (middle, horizontal black solid line in Fig. 6), and this trend and tendency will continue until 2050, affecting all Canadian forests. Therefore, one can expect significant ecological and environmental deterioration in addition to great human and economic losses in Canada in the next three decades.

The Bayesian multi-decadal model can explain the evolution of Canadian wildfires and the complex changes in the burned area. This power of explanation is especially relevant for the decrease of Canadian wildfires in the last two decades, which cannot be understood nor explained when the is strictly warming surface temperatures in Canada.

From 1930 to 1965, there was a very low activity phase of Canadian wildfires and low burned area (negative phase of the 60-year cycle). During this period, one can speculate that the AMO's positive phase causes more cloudiness, precipitation, and snow in Canada, so the vegetable fuel is wet, and the number of forest fires is low. It is during this negative phase that there were no reports of frequent wildfires nor any large, widespread wildfires, with one or two exceptions like the Chinchaga Firestorm of September 1950 engulfing 1.4 million ha of boreal forests of the Northern Alberta and British Columbia.

Then there is an increase in wildfires and the size of the burned area from 1965 to 1980 (positive and ascending phase of the 60-year oscillation). Thirdly, there is a stable high phase of both in the number of forest fires and in the area burned between 1980 and 1990. This stable interval takes place around the maximum of the 60-year cycle. After this decade of relative stability, a surprisingly decrease in wildfires and areas burned in Canada from 1990 till 2019 (descending phase and negative phase of the 60-year pattern). This overall positive phase of forest fires (1965-2000) coincides with the negative phase of the AMO, which causes less precipitation and snow, which is why fuel load has accumulated, and the number/frequency of forest fires is very high.

Large Canadian wildfires (see, for example Stocks et al., 2002) are reported during the positive phase of the trend of these fires but then again decrease substantially, or there was no major fire catastrophe during the last negative phase of the 60-year oscillation from the 2000s till present. Our Bayesian ML model predicts that this low fire frequency phase will probably last until 2030 (which coincides with the current positive phase of the AMO) and then a new high season of forest fires will begin (which will most likely coincide with the negative phase of the AMO), and the highest number of forest fires will peak at around 2040 -2045.

Also, even if the warm-dry hydroclimatic conditions for the 21st century might be conducive to increase fire frequency (Gaboriau et al., 2020). Nevertheless, this does not automatically mean a corresponding increase in wildfires' areal extent and intensity, especially if the more open landscape and particular vegetation type (i.e., conversion to more deciduous forests from coniferous type) prevail in Canada.

- Russian wildfires

The total area of forests in Russia is equivalent to 70% of the country's total land area. The meteorological conditions conducive for wildfires in Russia are: a) winters with little snow, b) a long period without rain, c) a high air temperature, and d) a low relative humidity. All these conditions are necessary, but they are not sufficient. Because for the appearance and development of a fire, two additional conditions are needed, such as the accumulation of vegetation fuel load and the presence of a fire/triggering source.

The main force of Russian forest fires is anthropogenic, and the second most important source is hydrometeorological and in the low latitude territory of Russia, the anthropic factor causes 98% of fires, and in high latitudes, it is electrical storms that cause 50% of forest fires. However, in some regions of northern Russia, 90% of forest fires can be caused by electrical storms.

For the third most important factor, the Russian wildfire patterns are caused by changes in the coupled atmosphere-ocean climate and weather system, and this aspect will be analyzed in this paper. In the case of Russia, the number of variables as input data for the MCW calculation is $N = 10$. The standardized time series, in addition to the number of yearly Russian forest fires, are ENSO, PDO, AMO, NAO, ACE, TSI, burned area, precipitation and surface temperature (top panel in Fig. 7). The historical record of Russian wildfires began in 1948.

The global time-averaged wavelet spectrum of the third MCW (left panel in Fig. 7) again shows both the decadal and new multi-decadal 30±5-year patterns. The first pattern is less than 95% of the confidence level and the second scale is more significant confidently established owing to the intrinsic shortness of the data records. The decadal and multi-decadal phase shows a complex nonlinear relationship among the total of ten variables studied. The multi-decadal cross-function of 30 years (blue curve in the bottom panel of Fig. 7) will be adopted to help offer a theoretical forecast for the next cycle of Russian wildfires, which will be compared with the Machine Learning predictive model. The global phase (right panel in Fig. 7) also reaffirms the complexity of the relationship between all ten co-variables.

We note the absence of annual patterns again because these are local factors in different Russian landscapes and forests and that seasonal climatic conditions also modulate these. We further note that the ENSO multi-annual patterns are also absent. Everything indicates that the effects of ENSO on wildfires are ultimately manifested on a more extended duration timescale, i.e., decadal and multi-decadal that are more intrinsically tied to the fuel-load climatic-ecological-environmental conditions and fire interactions

The modulation of ENSO by the TSI has been reported (Le Mouël et al., 2019; Weng, 2005; Douglass and Knox, 2015). Now, in addition to this

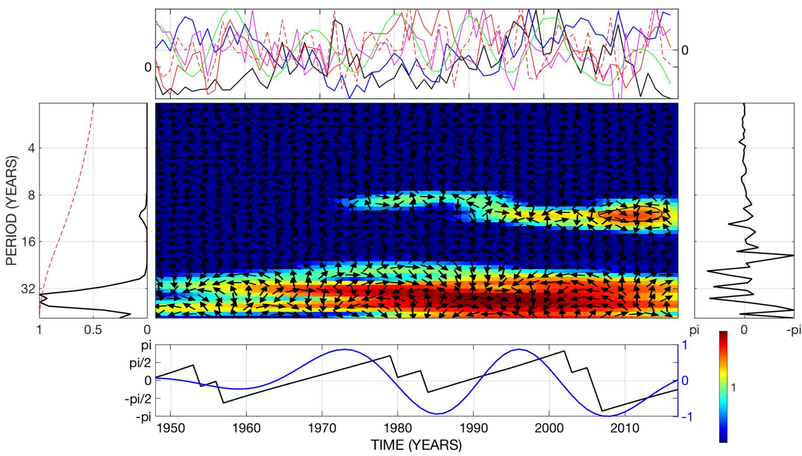

Figure 7: Time-frequency multi-cross wavelet from 1948 to 2019 between number of Russian wildfires, burned area, surface temperature, rainfall, snow cover, atmospheric-oceanic circulation and energy indices AMO, PDO, NAO, ENSO, ACE, and the external solar forcing factor TSI. The bottom panel shows the multi-decadal cross function at the significant timescale of 35-years (blue line) and the instantaneous phase relative for the same multi-decadal oscillation (black line). All other panels present similar information as described in Fig. 3 but for the Russian wildfire statistics.

pattern, its multi-decadal pattern on atmospheric and oceanographic circulations must be taken into account in order to offer a prediction of Russian wildfires.

We obtained a Bayesian Machine Learning (solid blue curve in Fig. 8) that describe the variability of Russian wildfires between 1948 and 2020 (*i.e.*, objective data), which are represented as a black line to the left of the vertical blue line of Fig. 8. It can be noted again that the historical annual-based data of these fires are well distributed around the Bayesian model. This model represents the multi-decadal frequency fluctuations of Russian wildfires.

It should be noted that during 1948-1965, 1975-1990 and 2010 and possibly until 2025, the precipitation in the Russian territory is well above its

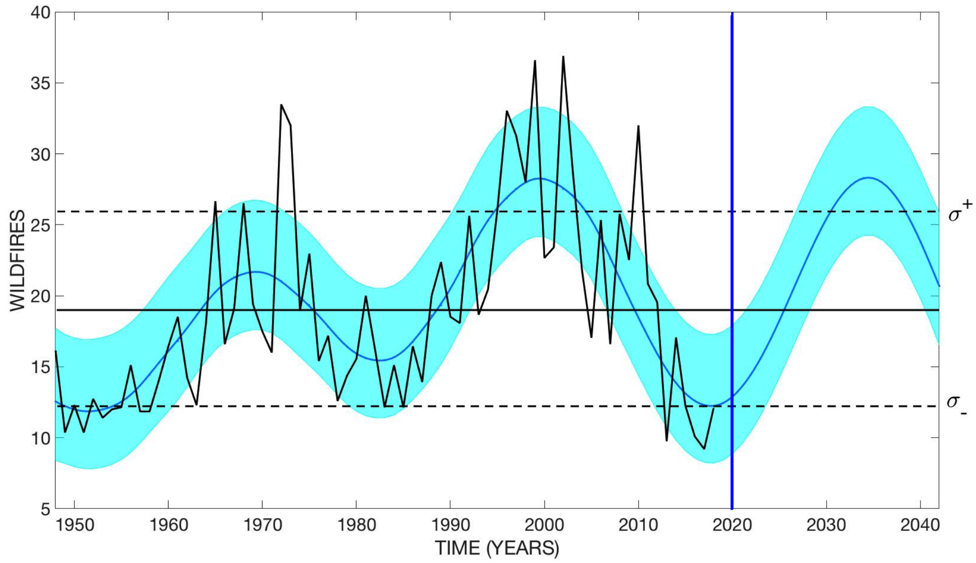

Figure 8: Annual frequency of wildfires in Russia from 1948 to 2019 (black line) compared with the Model's Machine Learning Bayesian inference (blue line). The horizontal solid black and dashed black lines are the mean fire frequency and its one standard deviation interannual statistics, respectively, for the objective data from 1948-2019 interval. The blue shaded area represents the 95% confidence intervals of the Bayesian ML model.

average value (hence an increase in fuel loads), and although El Nino events have been recorded, they have no clear and direct impacts on the increase in the number of fires or the increase in the burned area of Russian forests since the historical wildfire in Russia is at a minimum and are around σ^-. Notably, the highs (lows) of Russian wildfire activity occur in the negative (positive) phase of the 30-year PDO oscillation when rainfall and snow were well below (above) its historical average.

4. Discussion

Wildfires severely affect human society over the Northern Hemisphere, its safety, life, health, assets, economy, and properties, among other issues, every fire-active season without fail. They also affect the ecosystem and the environment. However, passive and emergency-based reactions and responses do not allow us to minimize the risks and vulnerability that Northern Hemisphere's society has endured and continue to suffer from wildfires. This is why we initiated this new framework to offer quantitative analyses and potential predictability of wildfire statistics in the Northern Hemisphere that may promise an operational transition from a quasi-passive response system like the one currently available to a disaster prevention system that may offer multi-years to even decade-long future horizons. Such a framework has a close tie to and need for remote sensing monitoring of the environment, as outlined in the further discussion below.

The fundamental difficulty of making any wildfire activity prediction for either near-term or long-term is the complex relationship between the variability of wildfires and global and regional climate changes, variations in the atmosphere-ocean circulation and transport modes, and even uncontrollable external factors. In addition, there are geographical and ecological settings to consider. If that were not enough, significant difficulties are analyzing the multi-factorial nature of the underlying deterministic and noisy relationships between wildfires and all relevant causative factors, including even arsons.

Faced with this complex situation and task, we show the powerful utility of the new techniques from Artificial Intelligence's Machine Learning. Machine Learning is a potent tool in analysing historical wildfire and its long-term prediction, especially when the available recorded baseline data are not enormously large. We develop a new methodology using various machine-learning algorithms and techniques to find the climatic patterns and ecologic conditions that induce high and low cycles of wildfires in each country of the Northern Hemisphere analyzed in this paperwork (the USA, Canada and Russia). Based on these intricate patterns of wildfires, we then strive to predict wildfires in longer-term.

Our newly developed methodology consists of differentiating from historical wildfire data the useful signals (i.e., deterministic patterns) and noises (i.e., stochastic fluctuations) to get a maximum signal to noise ratio. Find the climatic patterns that induce forest fires high and low cycles for each country in the North Hemisphere.

Once the patterns are identified, models to the average variability were trained, and these trained models can, in turn, explain more than 90% of the variations in the historical data. Trend models that can explain the complex variations in wildfires and the burned area were constructed. Subsequently, the models obtained are used to train the predictions of the following decadal or multidecadal cycles. We use MCW, Bayesian, and LS-SVM (i.e., described in the Method section below) in this process, but they can be replaced by any other algorithms of the user's preference. We examine any alleged relationship of the wildfire activity with climatic variables (including not only regional temperature and precipitation but also snow cover for the USA, Russia and Canada), several climatic circulations and moisture indices (i.e., ENSO, PDO, NAO, AMO and even Accumulation Cyclone Energy, ACE, indices), and external nudging factors like the Sun's irradiance activity cycles.

Specifically, we studied all the underlying co-factors responsible for the wildfire occurrence statistics for the USA, Canada, and Russia to find the climatic patterns that induce high and low wildfire cycles. We found that various combinations of local and regional climatic conditions, ocean-atmosphere circulation factors, and the external solar irradiance modulation factor can explain more than 90% of the wildfire frequency records.

The results obtained with Machine Learning from the analysis of the historical wildfires show that the decadal oscillation is present in the wildfires of all three Northern Hemisphere countries analyzed in this work. The difference is in the trend of these fires. While Canada has a multidecadal oscillation pattern of 60 ± 5 years, the USA and Russia have a 40 ± 5 and 30 ± 5 years modulation pattern. There are evidence for a decadal-like modulation of the wildfire frequency statistics in all three rather disparate geographical regions and ecological regimes. We interpret this set of empirical evidence to propose a quasi-decadal modulation of the regional fuel-load, hydro-climatic and wildfire conditions by the 11-yr Total Solar Irradiance (TSI) cycles. The relationship senses that precipitation is high during the 11-yr TSI high phase, the temperature is relatively cool or mild. Also, biomass fuel load build up with low natural fire frequency and tendency until the low TSI phase sets with dry and hot period with high fire frequency.

Finally, we proffer the forecast of emerging wildfire activity, adopting both the decadal and multidecadal oscillations identified over the USA, Canada and Russia using the Machine Learning training method. The results show that a new high cycle of forest fires has begun in each Northern Hemisphere's country (i.e., USA, Canada, and Russia) due to a combination of climatic variations (decadal and multidecadal) of the land-atmosphere- ocean system.

The Bayesian probabilistic forecasts for the USA, Canada and Russia also show a new high season of wildfires. It allows us to understand and be prepared for the future condition of the vulnerability, risk and danger of the boreal forests of the Northern Hemisphere associated with the impacts of climate change as well as anthropic activity.

In addition, these forecasts can be used to generate prevention actions in regions where the vegetation has shown greater vulnerability to forest fires. Evergreen and deciduous coniferous trees have been reported to be the most vulnerable cover to adverse atmospheric effects in the boreal areas of the Northern Hemisphere.

From the analysis of satellite data, we find that forest, agricultural, grassland, and scrub-type covers have the highest probability of ignition due to the frequency of fires associated with these land vegetation covers. Furthermore, the spread and permanence of fires during the months of July-December is associated with the accumulation and distribution of combustible materials in preceding events and with stressed vegetation due to the effect of climatic variability and human activity. In addition, the percentage of wildfires has increased between 2001 and 2018 by 8.9%, 46.5%, and 19.6% in the USA, Canada and Russia, respectively, for the main fire-affected vegetation category of tree cover, needle-leaved, evergreen vegetation covers

The increase in wildfires in non-forested vegetation cover is due to their greater capacity for ignition and fire propagation concerning boreal areas. In addition, its capacity for adaptation and resilience allows rapid regeneration and recovery, contributing to the accumulation of combustible materials mainly in transition zones (non-forested and forested land covers). These vegetation covers are located generally in lower elevation areas, where the frequency of fires is higher, limiting the forest regeneration process, contributing with new combustible materials and the colonization of vegetation vulnerable to the dominant climatic hazards for each study region.

5. Conclusions

The boreal forests of the USA, Canada and Russia are the most important carbon sinks. The percentage of wildfires affected vegetation category of the tree cover, needle-leaved, and evergreen vegetation cover has increased between 2001 and 2018 by 8.9%, 46.5%, and 19.6% in the USA, Canada and Russia, respectively. If the increase in wildfires continues in these countries, it could unbalance and overturn the Northern boreal forest's capacity as a carbon sink. This is why any capability to forecast wildfires reliably will be significant to minimize the risks and vulnerabilities of boreal forests of the

Northern Hemisphere.

We present a new methodology utilizing Machine Learning models with both purposes of developing models give insights into the complex relationship between the land-atmosphere-ocean system and Northern Hemisphere wildfires and the forecast of long-term wildfire. Our machine learning models show a new phase of high wildfire activity throughout the Northern Hemisphere has begun in 2020, created by decadal and multi-decadal variations of the coupled solar-land-atmosphere-ocean system.

Our ML model forecasts peak wildfires at around 2022±3, 2035±3, 2045±5 for the USA, Russia, and Canada, respectively. The new high wildfire activity phase will persist in the USA, Russia, and Canada until 2030, 2045, and 2055, respectively.

The results also indicate that a decadal oscillation occurs in wildfires of all three North Hemisphere countries with different varying patterns in each country. While the USA have another intrinsic oscillation of 40±5 years, Russia and Canada have oscillatory patterns of 30±5 and 60±5 years, respectively.

Acknowledgements

V.M. Velasco Herrera acknowledges the support from PAPIIT-IT102420 grant. W.S.'s effort is partially supported by CERES (http://ceres-science.com). G. Velasco Herrera acknowledges the support from "Marcos Mazari Menzer"-grant. We thank Drs. Michael and Ronan Connolly for proposing the statistical shuffling tests (which we tested but not shown in the paper) for checking the claimed detection of decadal and multidecadal co-variations in fire frequency statistics of USA, Canada and Mexico with other variables. Additional editing help from Jo Nova (http://joannenova.com.au/) is also deeply appreciated. Sergio Cerdeira-Estrada, Raúl Martell-Dubois, Jaime Valdéz, Laura Rosique-de la Cruz, Susana Perera-Valderrama, José Rosario López-Perea, Hansel Caballero Aragón, Rainer Ressl (2018-2020). Sistema de información y análisis marino-costero — Marine-coastal Information and Analysis System (SIMAR). CONABIO, México. Retrieved from https://simar.conabio.gob.mx

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to VMVH (**vmv@igeofisica.unam.mx**)

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Land Cover Classification System

Table A1: Classification of values from Land Cover Classification System (LCCS) to Land Cover (LC)

LCCS codes	Types of land covers
10	Cropland rainfed
20	Cropland irrigated or post-flooding
30	Mosaic cropland / natural vegetation (Tree, shrub, herbaceous cover)
40	Mosaic natural vegetation (Tree, shrub, herbaceous cover) / cropland
50	Tree cover, broadleaved, evergreen
60	Tree cover, broadleaved, deciduous, closed to open (>15%)
70	Tree cover, needleleaved, evergreen, closed to open (>15%)
80	Tree cover, needleleaved, deciduous, closed to open (>15%)
90	Tree cover, mixed leaf type (broadleaved and needleleaved)
100	Mosaic T and shrub / herbaceous cover
110	Mosaic herbaceous cover / T and shrub
120	Shrubland
130	Grassland
140	Lichens and mosses
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
160	Tree cover, flooded, fresh or brakish water
170	Tree cover, flooded, saline water
180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water
190	Urban areas
200	Bare areas
210	Water bodies
220	Permanent snow and ice

References

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., van der Werf, G. R., Pyne, S. J., 2009. Fire in the earth system. Science 32, 481–484.

Buduma, N., Locascio, N., 2017. Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. O'Reilly Media, Inc.

Cionco, R. G., Pavlov, D. A., 2018. Solar barycentric dynamics from a new solar-planetary ephemeris. Astronomy and Astrophysics 615, A153.

Cionco, R. G., Soon, W., 2015. A phenomenological study of the timing of solar activity minima of the last millennium through a physical modeling of the sun-planets interaction. New Astronomy 34, 164–171.

Daniau, A. L., Desprat, S., Aleman, J., Bremond, L., Davis, B., Fletcher, W., Marlon, J. R., Marquer, L., Montade, V., Morales-Molino, C., Naughton, F., Rius, D., Urrego, D. H., 2019. Terrestrial plant microfossils in palaeoenvironmental studies, pollen, microcharcoal and phytolith. towards a comprehensive understanding of vegetation, fire and climate changes over the past one million years. Revue de Micropaleontologie 63, 1–35.

Doerr, S. H., Santin, C., 2016. Global trends in wildfire and its impacts: perceptions versus realities in a changing world. Philosophical Transactions of the Royal Society B 371, 20150345.

Douglass, D. H., Knox, R. S., 2015. The sun is the climate pacemaker i. equatorial pacific ocean temperatures. Physics Letters A 379, 823–829.

Fedorov, V., 2018. Earth insolation and modern climate change. Fizmatlit.

Ferreira, L., Vega-Oliveros, D., Zhao, L., Cardoso, M., Macau, E., 2020. Global fire season severity analysis and forecasting. Computers & Geosciences 134, 104339.

Fletcher, T. L., Warden, L., Sinninghe Damsté, J. S.and Brown, K., Rybczynski, N., Gosse, J. C., Ballantyne, A. P., 2019. Evidence for fire in the pliocene arctic in response to amplified temperature. Climate of the Past 15, 1063–1081.

Fornacca, D., Ren, G., Xiao, W., 2017. Performance of three modis fire products (mcd45a1, mcd64a1, mcd14ml), and esa fire cci in a mountainous area of northwest yunnan, china, characterized by frequent small fires. Remote Sensing of Environment 9(11), 1131.

Gaboriau, D., Remy, C., Girardin, M., Asselin, H., Hély, C., Bergeron, Y., Ali, A., 2020. Temperature and fuel availability control fire size/severity in the boreal forest of central northwest territories, canada. Quaternary Science Reviews 250, 106697.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., Justice, C., 2018. The collection 6 modis burned area mapping algorithm and product. Remote Sensing of Environment 217, 72.

Giglio, L., Schroeder, W., Justice, C., 2016. The collection 6 modis active fire detection algorithm and fire products. Remote Sensing of Environment 178, 31.

Glover, K. C., Chaney, A., Kirby, M. E., Patterson, W. F., MacDonald, G. M., 2020. Southern california vegetation, wildfire, and erosion had non-linear responses to climate forcing during marine isotope stages 5-2 (120-15 ka). Paleoceanography and Palaeoclimatology 35, e2019PA003628.

Grinsted, A., Moore, J. C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical times series. Non-linear Processes in Geophysics 11, 561–566.

Hallett, D. J., Mathewes, R. W., Walker, R. C., 2003. A 1000-year record of forest fire, drought and lake-level change in southeastern british columbia, canada. Holocene 13, 751–76.

Ham, Y. G., Kim, J. H., Luo, J. J., 2019. Deep learning for multi-year enso forecasts. Nature 573, 568–572.

Hamilton, R., Stevenson, J., Li, B., Bijaksana, S., 2019. A 16,000-year record of climate, vegetation and fire from wallacean lowland tropical forests. Quaternary Science Reviews 224, 105929.

Han, Y., An, Z., Marlon, J. R., Bradley, R. S., Zhan, C., Arimoto, R., Sun, Y., Zhou, W., Wu, F., Wang, Q., Burr, G. S., Cao, J., 2020. Asian inland wildfires driven by glacial-interglacial climate change. Proceedings of the National Academy of Sciences of the USA 117 (10), 5184–5189.

Kappenberg, A., Lehndorff, E., Pickarski, N., Litt, T., Amelung, W., 2019. Solar controls of fire events during the past 600,000 years. Quaternary Science Reviews 208, 97–104.

Kennett, D. J., Kennett, J. P., West, G. J., Erlandson, J. M., Johnson, J. R., Hendy, I. L., West, A., Culleton, B. J., Jones, T. J., Stafford Jr., T. W., 2008. Wildfire and abrupt ecosystem disruption on california's northern channel islands at the allerod-younger dryas boundary (13.0-12.9 ka). Quaternary Science Reviews 27, 2530–2545.

Kitzberger, T., Brown, P. M., Heyerdahl, E. K.and Swetnam, T. W., T., V. T., 2007. Contingent pacific-atlantic ocean influence on multicentury wildfire synchrony over western united states. Proceedings of the National Academy of Sciences of the USA 104, 543–548.

Kubat, M., 2015. An introduction to machine learning. Springer `https://link.springer.com/book/10.1007/978-3-319-63913-0`.

Le Mouël, J. L., Lopes, F., Courtillot, V., 2019. A solar signature in many climate indices. Journal of Geophysical Research 124, 2600–2619.

Leal-Silva, M. C., Velasco Herrera, V. M., 2012. Solar forcing on the ice winter severity index in the western baltic region. Journal of Atmospheric and Solar-Terrestrial Physics 89, 98–109.

Liu, Z., Ballantyne, A., Cooper, L., 2019. Biophysical feedback of global forest fires on surface temperature. Nature Communications 10, 214.

Marriner, N., Kaniewski, D., Gambin, T., Gambin, B., Vanniere, B., Morhange, C., Djamali, M., Tachikawa, K., Robin, V., Rius, D., Bard, E., 2019. Fire as a motor of rapid environmental degradation during the earliest peopling of malta 7500 years ago. Quaternary Science Reviews 212, 199–205.

McCabe, G. J., Betancourt, J. L., Gray, S. T., Palecki, M. A., G., H. H., 2008. Associations of multi-decadal sea-surface temperature variability with us drought. Quaternary International 88, 31–40.

McCabe-Glynn, S., Johnson, K. R., Strong, C., Berkelhammer, M., Sinha, A., Cheng, H., Edwards, R. L., 2013. Variable north pacific influence on drought in southwestern north america since ad 854. Nature Geoscience 6, 617–621.

Melott, A. L., Thomas, B. C., 2019. From cosmic explosions to terrestrial fires? Journal of Geology 127, 475–481.

Moore, C. R., Brooks, M. J., Goodyear, A. C., Ferguson, T. A., Perrotti, A. G., Mitra, S., Listecki, A. M., King, B. C., Mallinson, D. J., Lane,

C. S., Kapp, J. D., West, A., Carlson, D. L., Wolbach, W. S., Them II, T. R., Harris, M. S., Pyne-O'Donnell, S., 2019. Sediment cores from white pond, south carolina, contain a platinum anomaly, pyrogenic carbon peak, and coprophilous spore decline at 12.8 ka. Scientific Reports 9, 15121.

Nanavati, W. P., Whitlock, C., Iglesias, V., de Porras, M. E., 2019. Post-glacial vegetation, fire and climate history along the eastern andes, argentina and chile (lat. 41-55^{deg}s). Quaternary Science Reviews 207, 145–160.

Pausas, J. G., Keeley, J. E., 2019. Wildfires as an ecosystem service. Frontiers in Ecology and the Environment 17, 289–295.

Pino, M., Abarzúa, A. M., Astorga, G. e. a., 2019. Sedimentary record from patagonia, southern chile supports cosmic-impact triggering of biomass burning, climate change, and megafaunal extinctions at 12.8 ka. Scientific Reports 9, 4413.

Restaino, C., Young, D. J. N., Estes, B., Gross, S., Wuenschel, A., Meyer, M., Safford, H., 2019. Forest structure and climate mediate drought-induced tree mortality. Ecological Applications 29, e01902.

Rimmer, S. M., Hawkins, S. J., Scott, A. C., Cressler III, W. L., 2015. The rise of fire: Fossil charcoal in late devonian marine shales as an indicator of expanding terrestrial ecosystems, fire, and atmospheric change. American Journal of Science 315, 713–733.

Rosan, T. M., Aragao, L. E. O. C., Oliveras, I., Philips, O. L., Malhi, Y., Gloor, E., Wagner, F. H., 2019. Extensive 21st-century woody encroachment in south america's savanna. Geophysical Research Letters 46, 2019GL082327.

Schreuder, L. T., Hopmans, E. C., Castaneda, I. S., Schefuß E., Mulitza, S., Sinninghe Damste, J. S., Schouten, S., 2019. Late quaternary biomass burning in the northwest africa and interactions with climate, vegetation, and humans. Paleoceanography and Palaeoclimatology 34, 153–163.

Scott, A. C., 2000. The pre-quaternary history of fire. Palaeogeography, Palaeoclimatology, Palaeoecology 164, 281–329.

Sejnowski, T. J., 2020. The unreasonable effectiveness of deep learning in artificial intelligence. Proceedings of the National Academy of Sciences of the USA 117.

Shen, H., Tao, S., Chen, Y., Odman, M., Zou, Y., Huang, Y., Chen, H., Zhong, Q., Zhang, Y., Chen, Y., Shu, S., Lin, N., Zhuo, S., Li, B., Wang, X., Liu, W., Liu, J., Pavur, G. K., Russell, A. G., 2019. Global fire forecasts using both large-scale climate indices and local meteorological parameters. Global Biogeochemical Cycles 33, 1129–1145.

Soon, W., 2009. Solar arctic-mediated climate variation on multidecadal to centennial timescales: Empirical evidence, mechanistic explanation, and testable consequences. Physical Geography 30, 144–184.

Soon, W., Connolly, R., Connolly, M., 2015. Re-evaluating the role of solar variability on northern hemisphere temperature trends since the 19th century. Earth-Science Reviews 150, 409–445.

Soon, W., Dutta, K., Legates, D. R., Velasco, V., Zhang, W., 2011. Variation in surface air temperature of china during the 20th century. Journal of Atmospheric and Solar-Terrestrial Physics 73, 2331–2344.

Soon, W., Velasco Herrera, V. M., Cionco, R. G., Qiu, S., Baliunas, S., Egeland, R., Henry, G., 2019. Co-variations of chromospheric and photometric variability of the young sun analog hd 30495: Evidence for and interpretation of mid-term periodicities. Monthly Notices of the Royal Astronomical Society 483, 2748–2757.

Soon, W., Velasco Herrera, V. M., Selvaraj, K., Traversi, R., Usoskin, I., Chen, C. T. A., Lou, J. Y., Kao, S. J., Carter, R. M., Pipin, V., Severi, M., Becagli, S., 2014. A review of holocene solar-linked climatic variations on centennial to millennial timescales: Physical processes, interpretative frameworks and a new multiple cross-wavelet transform algorithm. Earth-Science Reviews 134, 1–15.

Spessa, A., Field, R., Pappenberger, F., Langner, A., Englhart, S., Weber, U., Moore, J., 2015. Seasonal forecasting of fire over kalimantan, indonesia. Natural Hazards and Earth System Sciences 15(3), 429–442.

Stocks, B. J., Mason, J. A., Todd, J. B., Bosch, E. M., Wotton, B. M., Amiro,

B. D., Flannigan, M. D., Hirsch, K. G., Logan, K. A., Martell, D. L., Skinner, W. R., 2002. Large forest fires in canada, 1959–1997. Journal of Geophysical Research: Atmospheres 107 (D1), FFR 5–1–FFR 5–12.

Suykens, J. A. K., Gestel, T. V., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least-squares support vector machines. World Scientific Publishing Co., Pte. Ltd.

Torrence, C., Compo, G., 1998. A practical guide to wavelet analysis. Bulletin of American Meteorological Society 79, 61–78.

Turco, M., Jerez, S., Doblas-Reyes, F. J., AghaKouchak, A., Llasat, M. C., Provenzale, A., 2018. Skilful forecasting of global fire activity using seasonal climate predictions. Nature Communications 9, 2718.

Vapnik, V., 1998. Statistical learning theory. John Wiley and Sons, New York.

Velasco Herrera, V. M., Soon, W., Velasco Herrera, G., Traversi, R., Horiuchi, K., 2017. Generalization of the cross-wavelet function. New Astronomy 56, 86–93.

Weng, H., 2005. The influence of the 11 yr solar cycle on the interannual-centennial climate variability. Journal of Atmospheric and Solar-Terrestrial Physics 67, 793–805.

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., Lettenmaier, D. P., 2019. Observed impacts of anthropogenic climate change on wildfire in california. Earth's Future 7, 2019EF001210.

Wolbach, W. S., Ballard, J. P., Mayewski, P. A., Parnell, A. C., Cahill, N., Adedeji, V., Bunch, T. E., Domínguez-Vázquez, G., Erlandson, J. M., Firestone, R. B., French, T. A., Howard, G., Israde-Alcántara, I., Johnson, J. R., Kimbel, D., Kinzie, C. R., Kurbatov, A., Kletetschka, G., LeCompte, M. A., Mahaney, W. C., Melott, A. L., Mitra, S., Maiorana-Boutilier, A., Moore, C. R., Napier, W. M., Parlier, J., Tankersley, K. B., Thomas, B. C., Wittke, J. H., West, A., Kennett, J. P., 2018. Extraordinary biomass-burning episode and impact winter triggered by the younger dryas cosmic impact ∼12,800 years ago. 1. ice cores and glaciers. Journal of Geology 126, 165–184.

Zhang, Z., Wang, C., Lv, D., Hay, W. W., Wang, T., Cao, S., 2020. Precession-scale climate forcing of peatland wildfires during the early middle jurassic greenhouse period. Global and Planetary Change 184, 103051.

Zubkova, M., Boschetti, L., Abatzoglou, J. T., Giglio, L., 2019. Changes in fire activity in africa from 2002 to 2016 and their potential drivers. Geophysical Research Letters 46, 7643–7653.