

1 **Improving the Representation of Fire Disturbance in**
2 **Dynamic Vegetation Models by Assimilating Satellite**
3 **Data: a Case Study over the Arctic**

4 **E.P Kantzas^{1,2}, S. Quegan¹ and M. Lomas¹**

5 [1]{School of Mathematics and Statistics, University of Sheffield, Hicks Building,
6 Hounsfield Rd, Sheffield S37RH, UK }

7 [2]{Nansen International Environmental and Remote Sensing Centre, Vasilievsky
8 Island, 199034, St. Petersburg, Russia }

9 Correspondence to: E.P Kantzas (e.kantzas@sheffield.ac.uk)

10

11 **Abstract**

12 Fire provides an impulsive and stochastic pathway for carbon from the terrestrial
13 biosphere to enter the atmosphere. Despite fire emissions being of similar magnitude
14 to Net Ecosystem Exchange in many biomes, even the most complex Dynamic
15 Vegetation Models (DVMs) embedded in General Circulation Models contain poor
16 representations of fire behaviour and dynamics, such as propagation and distribution
17 of fire sizes. A model-independent methodology is developed which addresses this
18 issue. Its focus is on the Arctic where fire is linked to permafrost dynamics and on
19 occasion can release great amounts of carbon from carbon-rich organic soils.
20 Connected Component Labeling is used to identify individual fire events across
21 Canada and Russia from daily, low-resolution burned area satellite products, and the
22 obtained fire size probability distributions are validated against historical data. This
23 allows the creation of a fire database holding information on area burned and
24 temporal evolution of fires in space and time. A method of assimilating the statistical
25 distribution of fire area into a DVM whilst maintaining its Fire Return Interval is then
26 described. The algorithm imposes a regional scale spatially dependent fire regime on
27 a sub-scale spatially independent model; the fire regime is described by large scale
28 statistical distributions of fire intensity and spatial extent, and the temporal dynamics
29 (fire return intervals) are determined locally. This permits DVMs to estimate many
30 aspects of post-fire dynamics that cannot occur under their current representations of

1 fire, as is illustrated by considering the modeled evolution of land cover, biomass and
2 Net Ecosystem Exchange after a fire.

3

4 **1. Introduction**

5 Despite the high uncertainties in estimates of global biomass stocks, analysis of
6 carbon stock studies (Keith et al., 2009) shows that the boreal regions hold
7 considerable biomass per unit area, arising not only from the extent of boreal
8 ecosystems but also from the observed greening of the Arctic (Jia et al., 2003; Xu et
9 al., 2013); indeed, shrub communities have expanded during the 20th century (Sturm
10 et al., 2001), while tundra biomass has increased by almost 20% over the past three
11 decades (Epstein et al., 2012). Furthermore, due to low temperatures and consequent
12 low decomposition rates, enormous stocks of soil carbon exist in the Arctic (Ping et
13 al., 2008; Schepaschenko et al., 2013), some locked in and under permafrost, with the
14 soil organic carbon in the circumpolar permafrost region accounting for
15 approximately 50% of the global soil organic carbon pool (Tarnocai et al., 2009).

16 Changes and feedbacks in the fluxes of carbon between the land surface and the
17 atmosphere are of utmost importance in the context of global warming. The need to
18 gain quantitative understanding of such processes has encouraged the use of Dynamic
19 Vegetation Models (DVMs), often coupled to atmospheric models. DVMs simulate a
20 host of mechanisms linked to the terrestrial carbon and water cycles, with the aim of
21 reproducing the present status of the terrestrial carbon pools and fluxes and predicting
22 their trends. An influential and dynamic pathway by which terrestrial carbon enters
23 the atmosphere is through burning of vegetation and carbon-rich soils; its
24 implementation in DVMs is investigated in this study.

25 Key characteristics of fires include their inter-annual variability, size distribution and
26 intensity, which differ between regions of the Arctic (Wooster and Zhang, 2004).
27 Fieldwork and Earth Observation (EO) data show that larger fires contribute most to
28 the total area burned, despite being much rarer. From 1959-1999, out of the 1,000-
29 14,000 forest fires that occur every year in Canadian forests, only 3% exceeded 2 km²
30 in area but these accounted for 97% of the total area burned (Stocks et al., 1998).
31 Such fires usually last for several days or even weeks, and extend over large areas and

1 across biomes; in their central areas nearly all the vegetation is burnt, with reducing
2 degree of burn towards the fire scar edges.

3 Advances in EO sensors have contributed greatly to the information available on a
4 variety of fire characteristics, beginning with the detection of gas flares from oilfields
5 in images obtained from the AVHRR sensor on board the TIROS-N satellite (Matson
6 and Dozier, 1981). In the 21st century, images acquired from the MODIS instrument
7 are routinely used to identify fire scars at 500 m resolution (Roy et al., 2008) and
8 examine global trends in burned area (Giglio et al., 2010); by measuring thermal
9 anomalies the ATSR instrument can locate active fires and construct time series
10 monitoring their annual evolution (Arino et al., 2012); and measurements of fire
11 radiative power from geostationary and polar orbiting EO sensors allow the amount of
12 biomass lost to fires to be estimated (Wooster et al., 2012; Wooster and Zhang, 2004).

13 Nonetheless, the representation of fire in most DVMs does not utilize EO information
14 and fails to capture many of the key fire characteristics (Kantzas et al., 2013). A
15 typical DVM will estimate a fraction of area burned for each grid-cell based on
16 climate data (e.g. temperature and precipitation), vegetation characteristics (e.g. plant-
17 specific fire resistance) and other simulated variables (e.g. litter moisture), so the
18 outputs are deterministic and without any random component. Nevertheless, it is well
19 established that the size distribution of forest fires at continental scale follows the law
20 of small numbers and can be simulated stochastically with a Poisson model
21 parameterized with climate data (Jiang et al., 2012; Podur et al., 2010; Wiitala, 1999).
22 This heavily skewed distribution assigns high probability to small fires and lower
23 probability to bigger ones.

24 Most DVMs are unable to simulate large fires that occupy significant fractions of a
25 model grid-cell (which for a typical DVM resolution of 0.5° has dimensions of around
26 56 km by 28 km at 60° N). In addition, most DVMs are essentially point-based, with
27 no interaction between neighbouring grid-cells, so cannot simulate the propagation of
28 fire across several grid-cells. Instead, each grid-cell is assigned a small amount of fire
29 each year, with very little inter-annual variability (Kloster et al., 2010; Li et al., 2014;
30 Prentice et al., 2011; Thonicke et al., 2010; Thonicke et al., 2001). As an example, for
31 a typical year over the Arctic, in the LPJ-WM model (Wania et al., 2009a, b), which
32 is a version of the influential LPJ DVM (Sitch et al., 2003) tailored for high-latitudes,

1 the average fractional area burned per grid-cell is 0.3% with a variance of 0.045%,
2 and it rarely exceeds 1% in any grid-cell. This weakness in fire representation is
3 hidden when only the average fraction of area burned over a long period (whose
4 reciprocal is the Fire Return Interval or FRI) is reported. For example, if a DVM
5 represents fire by burning 0.5% of each grid-cell every year, the FRI will be 200 years,
6 but this completely fails to capture the highly episodic nature of boreal fires, in which
7 severe fire years may give emissions many times greater than the average (van der
8 Werf et al., 2010). In reality, observed FRI data from many small and some rarely
9 occurring big fires generally have a significantly higher variance than that produced
10 by DVMs. Correct simulation of the FRI is insufficient for DVMs to make accurate
11 predictions under changing climate scenarios.

12 The treatment of fire in DVMs also prevents them from capturing post-disturbance
13 dynamics (e.g. permafrost thawing and carbon fluxes) from large fires which remove
14 a considerable fraction of vegetation and soil carbon (Kantzas et al., 2013), like the
15 unprecedented 2007 Anaktuvuk River fire (Mack et al., 2011) which accounted for
16 almost 40% of the area burned in the entire state of Alaska for that year. Viereck and
17 Dyrness (1979) showed that the permafrost active layer depth increased from 40 cm
18 to 140 cm seven years after a fire disturbance in a black spruce Alaskan ecosystem
19 and Viereck (1983) showed that these effects can last for up to 30 years. Post-fire
20 carbon fluxes exhibit complicated dynamics (Amiro et al., 2003; Amiro et al., 2006),
21 with consequences for the extent to which vegetation recovery eventually turns a
22 region burned in a large fire from a carbon source into a sink, and how long, if ever, it
23 takes for carbon stocks to return to previous levels under a changing climate (Amiro
24 et al., 2001a).

25 The amount of litter removed in a fire is a key quantity controlling post-disturbance
26 permafrost degradation (Harden et al., 2006; Yoshikawa et al., 2002) while the water
27 cycle is also affected by large fires, and regional models have showed that changing
28 fire regimes cause changes in evapotranspiration in boreal forests (Bond-Lamberty et
29 al., 2009). Field data show that a substantial loss of canopy will decrease
30 evapotranspiration (Amiro et al., 2006) and canopy interception and consequently
31 increase groundwater recharge (Clark et al., 2012), but vegetation succession would
32 further complicate water dynamics, especially when forests stands are succeeded by
33 grass/shrubs for a number of years or indefinitely (Dore et al., 2010; Dore et al., 2012).

1 The DVMs would also be better coupled with atmospheric models and provide a more
2 realistic gas exchange interface if their simulations were capable of producing large
3 fires, with effects ranging from realistically simulating the carbon and trace gases
4 fluxes of big disturbances (van der Werf et al., 2010) to how smoke affects cloud
5 formation over the boreal forests (Sassen and Khvorostyanov, 2008) and the Amazon
6 (Koren et al., 2004).

7 Hence there are pressing reasons to improve the fire representation in the DVMs, but
8 these models are complex, involve highly coupled internal processes, operate on a
9 grid-cell basis, and are often embedded in climate models. In addition, significant
10 resources have been spent to calibrate fire processes so that the FRI compares well (in
11 some cases) with data (Prentice et al., 2011; Thonicke et al., 2010). Hence it is
12 desirable to keep model restructuring to a minimum and preserve its estimate of FRI,
13 while ensuring that fire characteristics, such as structure and size distribution, are
14 consistent with observational data.

15 The first step towards this goal is to obtain realistic statistical information on fire at
16 spatial scales appropriate to the models, i.e. 0.5° - 1° , for example the number of fires
17 per year, their size distribution and their spatial characteristics. Currently, historical
18 information on wildfires in the Arctic, such as their number, area burned and
19 boundaries, is compiled in databases by fire agencies in Canada (Canadian National
20 Fire Database) and Alaska (Alaska Interagency Coordination Centre); these consist
21 mostly of ground observations supplemented by EO data. Due to the remoteness of
22 much of the boreal zone, there are large data gaps, and similar data do not exist for the
23 much larger area of the Russian Arctic. In Section 2.1 we show how readily-available
24 image analysis tools, specifically Connected Component Labeling, can be employed
25 to identify individual fires in EO burned area data and extract the information needed
26 to characterize fires in the Arctic statistically. We then exploit this information in
27 Section 2.2 to develop a model-independent methodology for creating fires with a
28 realistic size distribution in DVMs while maintaining their FRI and involving little
29 model restructuring. In Section 3 we verify both methodological approaches and
30 demonstrate some of the consequences for post-fire dynamics, while in Section 4 we
31 discuss the limitations of the approach and possible ways to address them.

32 **2. Methodology**

1 2.1. Connected Component Labeling

2 Connected component labeling (CCL) (Gonzalez et al., 2003) or “blob detection” in
3 the context of image processing is a method where unique clusters in a binary image
4 are identified based on the connectivity of their sides and/or edges. In two dimensions,
5 two categories exist: 4-connected and 8-connected. In 4-connected labeling, each
6 pixel with coordinates (x, y) can be connected to those pixels with which it shares an
7 edge, i.e. the pixels with coordinates $(x \pm 1, y)$ and $(x, y \pm 1)$. In 8-connected labeling,
8 pixels with a common vertex are also included, so there are extra possible connections
9 to the pixels at positions $(x \pm 1, y \pm 1)$. Thus in a binary image, CCL would *label* or
10 *cluster* connected blobs of 1s against a background of 0s. Two dimensional CCL has
11 numerous applications in image analysis, and has been used for clustering pixels in
12 fire scars in single images (Koltunov et al., 2012; Morissette et al., 2005). CCL can
13 also be applied in three dimensions, where the third dimension can be time, and we
14 exploit this capability to determine the growth of fire scars in sequences of daily EO
15 images of burned area. For each image, pixels identified as burned are assigned the
16 value 1 and the rest are given the value 0. Additionally each image and its pixels are
17 labeled by the associated day of the year, t , to yield a 3-dimensional dataset (x, y, t) on
18 which we apply the CCL algorithm.

19 Three-dimensional CCL has 6, 18 and 26-connected categories, defined respectively
20 at a given voxel by those voxels having a common face, plus those with a common
21 edge, plus those with a common vertex; the first 2 categories are depicted in Fig. 1.
22 Assuming the accuracy of the underlying daily burned area images, CCL should be
23 able to track the progress of a particular fire from its ignition, through its temporal and
24 spatial propagation to its extinction, by following the connections between burned
25 pixels. As fire scars are continuous in both space and time, individual fires will be
26 labeled and subsequently categorized based on burned area. It should be noted that the
27 labelling of an individual fire may depend on the spatial resolution of the EO sensor,
28 since what is seen as a single burned pixel at lower resolution may in fact be resolved
29 as several fires when imaged at higher resolution. However, this does not cause
30 problems when assimilating the data, as the model used will have the same spatial
31 resolution as the fire database created.

1 In principle, the 6-connected variety of CCL should be sufficient to capture fire
2 spread as a fire could not propagate diagonally in space without affecting the adjacent
3 pixels. For example, a fire at (x, y, t) propagating to $(x + 1, y + 1, t)$ would most likely
4 affect $(x + 1, y, t)$ and/or $(x, y + 1, t)$. However, in some cases the fire may propagate
5 diagonally with no detectable effect on the adjacent pixels, for example because of
6 sensor detection sensitivity, so 6-connected CCL would detect it as two independent
7 fires instead of a single event. It is also possible that two fires occurred on the same
8 day in diagonal grid-cells with independent ignition sources, whether natural or
9 anthropogenic. Weather conditions conducive to lightning can cover large areas and
10 lead to lightning igniting more than one fire in a wide front, so whether these fires are
11 independent smaller fires or a single larger one is a matter of interpretation. Hence we
12 applied both the 6 and 18-connected CCL algorithm and compared the results with
13 available data on fire statistics in order to determine which was more appropriate.

14 **2.1.1. Datasets**

15 The CCL algorithm was applied to the latest version (v.4.0) of the influential Global
16 Fire Emissions Database (GFED4) (Giglio et al., 2013). This is based on the
17 algorithm of Giglio et al. (2009) and provides two products: burned area (GFED-BA),
18 which gives the area burned within each grid-cell, and fire emissions, which gives
19 fire-induced emissions of various chemical species, such as CO_2 , CH_4 and NO_x (van
20 der Werf et al., 2010). For the period used in this study, from the mid-2000s to the
21 present day, the GFED-BA is derived daily from the MODIS MCD64A1 500 m
22 burned area product (Roy et al., 2008), which is based on changes in reflectance in the
23 visible channels of MODIS, but the GFED-BA also takes into account information on
24 active fire counts (Giglio et al., 2009). It is not offered at the MODIS 500 m
25 resolution but instead is aggregated to a resolution of 0.25° to facilitate interfacing the
26 fire data to biochemical and atmospheric models which run at such resolutions
27 (Castellanos et al., 2014; Kaiser et al., 2012; Valentini et al., 2014).

28 The Canadian Large Fire Database (CLFD) (Stocks et al., 2002) offers the best tool to
29 test the outputs from the CCL analysis. It reports on forest fires greater than 2 km^2 in
30 extent occurring in Canada from 1959-1999, including their date, location and size,
31 together with metadata such as cause of ignition, when available. The CLFD has been
32 used extensively in various contexts, such as investigating temporal trends in burned

1 area (Krezek-Hanes et al., 2011), evaluating fire emissions (Amiro et al., 2004; Amiro
2 et al., 2001b) and modeling fire frequency (Jiang et al., 2012).

3 In order to evaluate the CCL algorithm against the CLFD, the 0.25° GFED4 grid-cells
4 that contain forest in Canada must first be identified. Instead of utilizing a land cover
5 product, which would add unnecessary uncertainty, we built the forest map by
6 combining the CLFD and GFED4 data. To do this, we first applied the CCL algorithm
7 to the GFED4 data and assigned the value 1 to a grid-cell if it also contained a fire
8 record in the CLFD. Clearly this would omit forest grid-cells where the CLFD did
9 record any fire over its 40-year period, so to generate a forest mask the set of
10 identified pixels was morphologically closed. This assigns 1s to grid-cells in close
11 proximity to or surrounded by grid-cells already assigned the value 1. All other pixels
12 were considered as non-forest and assigned the value 0.

13 We also applied the CCL algorithm over Russia, but here used the GlobCover 2000
14 land cover map (Arino et al., 2008) to distinguish forest from non-forest. The area of
15 forest-related classes within each grid-cell was aggregated and if this exceeded 50%
16 the grid-cell was assigned as forest, otherwise as non-forest.

17 **2.2. Assimilating CCL fire products into a Dynamic Vegetation Model**

18 Applying CCL to the daily GFED4 burned area images from 2001-2012 allows the
19 creation of a database of individual fire events that includes their geographical
20 location, daily propagation, fire size and geometry, i.e. how many grid-cells were
21 affected and the fraction of each that was burned. We now give details of a
22 methodology that assimilates this information to produce a realistic fire regime in a
23 DVM whilst maintaining its internally simulated FRI. The algorithm can be applied to
24 any sub-grid of pixels whose aggregate geographical representation is considered to
25 have a spatially independent fire regime in terms of size and intensity. Here it is
26 applied separately to Canada and Russia, which are considered to have different fire
27 regimes.

28 1) A DVM calculates the annual fraction of area burned in year y , $\mathbf{BA}(lat, long, y)$
29 for each grid-cell, where lat and $long$ denote latitude-longitude. As described
30 in the Introduction, in most current DVMs only 0.1% - 5% of each grid-cell
31 burns annually. Each year we accumulate this fractional burned area into a

1 new cumulative array, **BAC**, which gives the total fractional area per grid-cell
 2 burned after y years, and is defined as:

$$3 \quad \mathbf{BAC}(lat, long, y) = \sum_{n=1}^y \mathbf{BA}(lat, long, n). \quad (1)$$

4 2) For each year we integrate $\mathbf{BA}(lat, long, y)$ over its spatial dimensions to give
 5 the aggregated fraction of area burned in year y ,

$$6 \quad int_f(y) = \iint_{lat/lon} \mathbf{BA}(lat, long, y). \quad (2)$$

7 As an example, in LPJ-WM the value of int_f for a representative year is
 8 approximately 28.0 for Canada and 47.0 for Russia. Since the numbers of LPJ
 9 0.5° grid-cells in the two countries are approximately 8000 and 12500
 10 respectively, the model burns an average fraction of 0.35% per grid-cell for
 11 Canada and 0.375% for Russia; in both cases, northern Arctic grid-cells
 12 significantly reduce the overall average fraction burned.

13 3) Using CCL-6, we created a database [**CCL-6**] which, as explained above,
 14 labels all grid-cells belonging to a single fire and records the fraction burned
 15 in each of these grid-cells. For each fire, we sum these fractional areas. For the
 16 majority of the fires, e.g. a small fire over a single grid cell, the summation
 17 will yield a percentage close to 0.1%, but for larger fires that spread over
 18 multiple grid-cells, e.g. 15 grid cells with an average 10% burn, the
 19 summation can exceed 100%. We then average these summations for all fires
 20 in the database to give μ_{fire} . For Canada, μ_{fire} is 1.23% and for Russia it is
 21 0.76%.

22 4) We define the total number of fires in a specific year y to be

$$23 \quad no_{fires}(y) = \frac{int_f(y)}{\mu_{fire}}, \quad (3)$$

24 which amounts to approximately about 2000 fires for Canada and 6000 for
 25 Russia, depending on year.

26 5) We then randomly select with replacement from the [**CCL-6**] database a
 27 number of fires occurring in year y equal to $no_{fires}(y)$. The total fraction of
 28 area burned will therefore be a normally distributed random variable with
 29 mean

$$30 \quad \mu_{fire} * no_{fires}(y) \quad (4)$$

1 and variance

$$2 \quad no_{fires}(y) * variance([CCL - 6]), \quad (5)$$

3 where $variance([CCL-6])$ is 1.02% and 2.69% for Canada and Russia
4 respectively. This process would cause the total fraction of area burned for
5 that year to be a random variable, but we wish to fix it to $int_f(y)$ which is the
6 fraction that the model wants to burn in year y (Step 2), so we normalize the
7 size of each fire so that

$$8 \quad int_f(y) = \mu_{fire} * no_{fires}(y). \quad (6)$$

9 6) Each fire selected from the **[CCL-6]** database is then overlaid on a randomly
10 selected subset of **BAC**($lat, long, y$) with the same spatial dimensions as the
11 fire, e.g. if the selected fire is extended over 3x1 grid-cells then a 3x1 grid-cell
12 area will be randomly selected from **BAC**. If each grid-cell in the **BAC**($lat,$
13 $long, y$) subset has an accumulated fractional area burned greater than or equal
14 to that of the corresponding-grid-cell in the selected fire, then the fire will be
15 accepted, i.e., considered to occur, and the fraction of each affected grid-cell
16 as given by the **[CCL-6]** overlay will be subtracted from **BAC**($lat, long, y$).
17 Otherwise, a new subset will be selected at random from **BAC** until a subset
18 capable of accommodating the fire is found.

19 7) The chance of finding a suitable location for a particular fire event decreases
20 with increasing fire intensity and extent, and such a location may not exist. In
21 the rare cases when this occurs the fire is forced to fit the location that came
22 closest to accommodating the fire and the deficit in **BAC** is taken from other
23 pixels to maintain the regional average of FRI. These cases occur on average
24 once every 2 years and the **BAC** deficit is approximately 50% of a grid-cell.

25 8) Since DVM calculations of FRI differ between regions according to climate
26 and vegetation, the subsets of **BA** with higher values will also have higher
27 values of **BAC** since the fractional burned area will accumulate there faster.
28 Hence these regions will be able to accept more fires and the random process
29 of selecting grid-cells will converge to produce a FRI equal to the reciprocal
30 of **BA**.

31 This methodology requires an initial run of the DVM to produce **BA** for each year.
32 These values are then fed into the above procedure to define the fires that are accepted
33 in the **BAC** array for that specific year. The grid-cells which experience burn and the

1 fraction burned are stored. The model is then rerun but with area burned read from the
2 outputs of the algorithm. Even though this requires two runs, the initial run to acquire
3 **BA** is not always needed. As long as the FRI of the model does not change
4 significantly, either the fires produced by a previous application of the algorithm can
5 be used or the algorithm can be run again with **BA** obtained from a previous run of
6 the model. In the latter case, and since the process is stochastic, a different set of fires
7 will be produced but the FRI will not change.

8 The LPJ-WM DVM used in this study calculates a daily fire probability for each grid
9 cell as a function of temperature and litter moisture (Thonicke et al., 2001). The fire
10 probability is then summed over the course of a year, from which the length of the fire
11 season and fraction of area burned per grid cell is derived; the values of the latter
12 populate the **BA** array. As DVMs are designed for global simulations, the built-in fire
13 probability function needs to produce FRIs that vary considerably (10-1000 years)
14 depending on the ecosystem. The sensitivity of fire probability to driving variables is
15 thus optimized to the geographic variability of climate, which is considerably higher
16 than the temporal variability over a grid cell; consequently the burned area of a grid-
17 cell remains largely constant. Our approach corrects this unrealistic behavior by
18 converting the fire simulations of DVMs into stochastic processes, the random
19 component of which provides temporal variability with the aid of data assimilation.

20 **3. Results**

21 **3.1. Applying Connected Component Labeling to the Canadian Large** 22 **Fire Database**

23 Following the approach of *Jiang et al.* (2012), the individual fires obtained by CCL-6
24 and CCL-18 for all four areas (Canadian forests and non-forests, Russian forests and
25 non-forests) were assigned to five categories according to fire size: (1) 2 km² to 10
26 km², (2) 10 km² to 30 km², (3) 30 km² to 100 km², (4) 100 km² to 500 km², (5) greater
27 than 500 km²; the aggregate of (1) to (5) is defined as a sixth category. We then
28 applied two non-parametric statistical tests to test the null hypothesis that the fire
29 sizes obtained from CCL and from the CLFD represent samples from the same
30 distribution. The two-sample Kolmogorov-Smirnov (KS) test uses a statistic that
31 quantifies the distance between the cumulative distribution functions of the two
32 samples; small values of this statistic indicate that the samples originate from the
33 same distribution. The two-sample Mann-Whitney-Wilcoxon (MWW) test examines

1 whether two independent samples originate from distributions with equal medians.
2 Both tests were performed at a 90% confidence interval with results shown in Fig. 2.

3 The best agreement was achieved between the CLFD and CCL-6 on Canadian forests.
4 Here, Categories 2, 3, 4 and 5 all passed the KS test while categories 1, 2, 3 and 5
5 passed the MWW test. Category 1 failed the KS test and Category 4 the MWW test.
6 The broad category 6 passed the MWW but not the KS test. When applied to
7 Canadian forests, CCL-18 detected 15% fewer fires than CCL-6 because the increased
8 number of connecting points in CCL-18 merged fires that CCL-6 characterized as
9 distinct. Nevertheless, the frequency distributions remained largely unchanged and
10 consequently the results of both statistical tests were identical for every category.

11 CCL identified fewer Canadian non-forest fires than forest fires as most of the non-
12 forest cover is in the smaller expanse of the Great Plains in the south and in the Arctic
13 north, where climate causes a much smaller fire occurrence frequency. The smaller
14 number of fires in non-forest grid-cells, in combination with the division into 5
15 categories and subdivision into 15 bins per category, reduces the size of the sample,
16 causing higher sample variance and a less smooth histogram than for forests (Fig. 2).
17 Nevertheless, categories 2 to 5 passed both statistical tests, but the 6th aggregated fire
18 category did not pass any of the tests. As seen in Fig. 2, this is because in Canada
19 non-forest fires produced by CCL have smaller sizes than for forest, which is also the
20 case when CCL is applied over Russia.

21 As no extensive, fire-related ground data are available for Russia, we compared the
22 results of the CCL algorithm over Russia against the CLFD. For forests, the MWW
23 test was passed for categories 1, 4 and 5 and the KS test for categories 3, 4 and 5.
24 Neither test was passed for the overall category 6 for forest or non-forest. As seen in
25 Fig. 2, this is because of the larger fraction of small forest fires compared to Canada.
26 Nevertheless, as noted earlier, the bigger fires contribute disproportionately to the
27 annual area burned and consequently are the most important to incorporate correctly
28 in the DVMs. Indeed, in the CLFD, fires over 30 km² accounted for 30.3% of the total
29 number of fires but contributed 91.2% of the area burned; similar results were
30 obtained with CCL-6 for Canadian forests (30.7% and 92.5%) and Russian forests
31 (19.8% and 89.6%) despite the non-overlapping time periods of the analysis.

1 Even though the fire size sample distributions for Canada and Russia were similar, the
2 statistical tests show that they did not originate from the same distribution. This may
3 be associated with the known differences in intensity between Canadian and Russian
4 forest fires (Harden et al., 2000; Wooster and Zhang, 2004). However, it could also be
5 a sampling artefact arising from the small number of years for which there are data in
6 GFED. For example, if the data available for Russia had covered more years with
7 large fires, the distribution of fire sizes would be shifted to the right and would more
8 closely match the distribution of the Canadian fires. Hence the lack of a database
9 analogous to the CLFD prevents safe conclusions to be drawn regarding the validity
10 of CCL results over this region.

11 The statistical tests show that the CCL algorithm produces a histogram of forest fire
12 sizes closely matching that from the CLFD, and it also produces a similar probability
13 function for Canadian non-forest, especially for the categories containing larger fires.
14 This agreement occurs despite the CLFD recording fires from 1959-1999 while the
15 GFED4 starts in 2001. This could indicate that, despite fluctuations in the number of
16 fires and area burned each year, their size distribution remains essentially unchanged,
17 an assumption implicit in the statistical tests performed.

18 To simplify the assimilation of the CCL database into a DVM we pooled forests and
19 non-forest fires as identified by CCL-6 together but maintained the distinction
20 between fires occurring in Canada and Russia.

21 **3.2. Fire disturbance simulations with assimilated CCL fire products**

22 To test whether the FRI is conserved between the initial and rerun version of the
23 model where CCL fires are utilized, we calculated **BA** from 1000 years of spin-up and
24 112 years of transient runs (1901-2012) of LPJ-WM driven by CRU 3.0 climatology
25 (Mitchell and Jones, 2005); we then ran the algorithm described in Section 2.2 for the
26 full 1112 years and produced a set of fires for each year for both Canada and Russia.
27 The FRI obtained using the new algorithm, referred to hereafter as a CCL run, closely
28 matched the FRI obtained from the original run, demonstrating that fire can be
29 included in a DVM in a way that retains the model structure and FRI, but is also
30 consistent with the size distribution of burned area observations (Fig. 3). Even though
31 the CCL run adds random spatial variability to the FRI, the average magnitude of FRI
32 remains largely unaffected over sub-regions of both Canada and Russia.

1 We investigated whether this variability in FRI is caused by the short spin-up time of
2 the DVM (1000 years) compared to the long FRI for the region (100-1000 years),
3 which may not allow enough time for the FRI to converge to the original model value.
4 However, even CCL runs with over 4000 years of spin-up failed to produce the
5 original spatially smooth FRI. Only after excluding the very large fires (Categories 4
6 & 5 in Fig. 2) from the CCL algorithm was the spatial variability reduced: an almost
7 exact match to the original FRI was then achieved. This seems to indicate that the
8 spatial variability arises from the limited number of large fires found by CCL, both
9 because they are rare and because GFED4 is derived from data covering only a
10 decade. The conditions imposed by the algorithm make it hard for them to be
11 accepted under comparison with the accumulated burned area array **BAC** (Section 2.2,
12 Step 6). As a result, these larger fires are frequently allowed to burn the same subsets
13 of grid-cells, which hinders production of a smooth FRI across the region.
14 Nevertheless, as Fig. 3 shows, the FRI produced by the CCL run does capture the FRI
15 of the original run in the sub-regions of the Arctic. Possible ways to reduce the
16 variability are discussed in Section 4.

17 **3.3. Post-Fire Dynamics**

18 Of greater importance is that the CCL run produces fire size characteristics consistent
19 with those derived from EO data. Fig. 4 demonstrates this by comparing the fraction
20 of burned area over a year of the transient run (2007) obtained from an original run of
21 LPJ-WM, a CCL run for the same year and GFED4. As noted in the Introduction, the
22 original LPJ-WM representation of fire (top) causes a very small fraction of most of
23 the grid-cells to burn, and the area burned (either per grid-cell or total) remains
24 largely unchanged in different years; such behaviour is common to many DVMs. In
25 contrast, using the CCL methodology (centre) gives rise to fires whose sizes cover the
26 entire range of burned areas, as shown in Fig. 2. Additionally, since the algorithm
27 accumulates grid-cell fire potential and then consumes it when one or more grid-cells
28 are chosen to accommodate a fire, regions are prevented from unrealistic behavior in
29 which big fires are separated by only a short time span. The long FRI in the Arctic
30 means that several decades need to pass after a big disturbance before a grid cell has
31 accumulated enough fire potential in order to experience another fire.

32 To investigate the benefits of this new ability to simulate large fires in DVMs, a fire
33 occurring in northwest Canada in simulation year 1910 is examined in Fig. 5. This

1 particular fire spread over 16 grid-cells; the fraction burned was approximately 80%
2 in the two central grid-cells and fell off towards the edges of the fire scar. Post-fire
3 competition amongst species in the CCL run gives rise to evolution of vegetation
4 cover that is consistent with field data (Dore et al., 2012). The fire occurred in a forest
5 region surrounded by herbaceous vegetation (Fig. 5 top, Fire Year -1). As expected, in
6 the year of the fire and that following (Fire Year 0 & +1), the dominant cover
7 switches to bare ground. By year +4, plant competition processes lead to the
8 vegetation becoming a mixture of grass and trees, with trees, as saplings, becoming
9 the dominant species by year +5. In contrast, biomass requires much more time to
10 recover. In Fire Year 0 and the years immediately after (Fig. 5, middle), the biomass
11 of the forest is similar in magnitude to that of the neighbouring grass grid-cells. As the
12 forest regenerates, biomass slowly recovers to pre-fire levels while the fire scar
13 remains visible in the model calculations even 50 years after the fire. Carbon fluxes
14 are expressed through the annual Net Ecosystem Exchange (NEE), which is the net
15 flux of carbon to the atmosphere from all possible pathways (Fig. 5, bottom). In Fire
16 Year 0, fire emissions turn the grid-cells into strong sources whose NEE is about an
17 order of magnitude off the scale used. Even though vegetation begins to recover, the
18 fire scar is initially (Fire Year +2) not a strong carbon sink, since the cover is mostly
19 grasses and saplings with limited carbon uptake rates. However, by Fire Year +10 it
20 has developed into a marked sink, even though the surrounding region for that
21 particular year happens to be a source. This indicates the value of this new approach
22 for simulating carbon dynamics in the Arctic boreal zone, since these effects cannot
23 occur in the original version of the DVM. However, as discussed in Section 4, it needs
24 further extension before being able to account for effects such as large scale inter-
25 annual variability.

26 **4. Discussion**

27 The new methodology for simulating fire disturbance in DVMs described in this
28 paper significantly improves on current model approaches by providing a
29 representation of fire sizes that is consistent with large scale satellite observations and
30 requires minimal modifications to model structure. Having realistic fire sizes offers
31 scope for investigating both post-fire carbon dynamics and effects on the water cycle
32 caused by large fires; this is of particular interest in the Arctic due to the presence of
33 permafrost, carbon-rich soils and an expected increase in fire activity caused by

1 changing climate (Balshi et al., 2009a; Balshi et al., 2009b; Krawchuk et al., 2009;
2 Stocks et al., 1998). However, our approach has certain limitations, which depend on
3 the DVM in which the methodology is applied, the region under study and the
4 methodology itself.

5 Taking advantage of this new capability requires DVMs with sufficiently rich process
6 representations; indeed, the lack of such a capability has meant there has been no
7 motivation for the DVMs to embody the process coupling that is set in train when
8 severe fires occur. This is particularly true as regards the connections between fire,
9 land cover and permafrost. It is also the reason why our post-fire analysis was
10 qualitative and restricted to examining whether a DVM has the capability to simulate
11 the expected ecosystem response following a fire; without all the necessary linked
12 processes in place, it is premature to attempt a quantitative comparison of carbon and
13 water fluxes between field data and simulations. For example, even though LPJ-WM
14 considers permafrost, the upper boundary value for soil heat transfer is the daily air
15 temperature provided by the climatology driver. Hence, even after a large fire that
16 removes most of the canopy, thermal conduction is unaffected, and no account is
17 taken of heating of the soil by incoming radiation. These shortcomings can be
18 alleviated in an ad hoc fashion by using an extinction equation parameterized by Leaf
19 Area Index to characterize temperature during canopy recovery but, as shown by
20 Kantzas et al. (2013), what is really required is a more sophisticated radiative/heat
21 transfer process. The JULES model (Best et al., 2011), for example, does consider
22 radiative transfer through the canopy and has a recently-added permafrost
23 representation which considers the thermal properties of organic soils (Chadburn et al.,
24 2015), but JULES does not contain a fire component.

25 The CCL algorithm captures all the temporal characteristics of a fire event (dates of
26 ignition and extinction and temporal evolution), but this information cannot be
27 assimilated in LPJ-WM because this DVM only calculates fire effects annually. This
28 significantly weakens the ability of LPJ-WM (and other DVMs with annual fire
29 accounting) to model, for example, post-fire permafrost dynamics, biomass burning
30 and litter/soil emissions, since the timing of a fire relative to summer defines its effect
31 on carbon pools and soil heat transfer. Nevertheless, and since the algorithm described
32 here is model-independent, a DVM with a daily fire step could be used, such as the
33 Community Land Model (CLM) (Kloster et al., 2010), although CLM does not

1 currently consider carbon soils or specific Arctic plant functional types like LPJ-WM.
2 However, the temporal characteristics of fires obtained by CCL could be assimilated
3 into a DVM by making the probability distribution of fire occurrence conditional on
4 day of the year or season, thus allowing post-fire dynamics to be studied at higher
5 temporal resolutions.

6 The long FRI in the Arctic means that the 12 years of data in the GFED4 daily
7 product is insufficient for adequate sampling of the rarer large fires. This can distort
8 the local occurrence statistics and give rise to spatial variability in the simulated FRI,
9 even though at larger scales the CCL runs agree well with the FRI produced by the
10 original model. The restricted number of larger fires means that the algorithm tries to
11 accommodate the same large fires over regular time intervals, which slightly alters the
12 FRI produced by the original model run. It seems likely that this spatial variability
13 would be reduced in regions and sub-regions with shorter FRI where acquiring
14 statistically representative data is less problematic. Available fire data would then
15 offer a more representative picture of the local fire regime. Alternatively, in such
16 regions, at each position an empirical distribution can be fitted to the histogram of fire
17 sizes identified by CCL, and this probability density function could be used for
18 sampling; this resolves problems associated with unoccupied bins in the histogram.

19 Burned area data from GFED4 over the Arctic reveals that in a given year fires tend to
20 cluster spatially (Fig.4, bottom), presumably because of fuel availability and
21 conditions that locally favour fire ignition and propagation, such as high temperatures
22 and winds, low precipitation and abundance of natural or anthropogenic ignition
23 sources. Such local weather conditions in turn give rise to inter-annual variability
24 (IAV) in burned area over larger regions, e.g. the increase observed in North America
25 during El Nino years (van der Werf et al., 2004). In contrast, the fires simulated by
26 assimilating the CCL algorithm in LPJ-WM (Fig. 4, centre), even though they have
27 the correct size distribution, do not appear in clusters and give a smaller IAV in
28 burned area than GFED4 data. This is because the assimilation preserves the original
29 annual burned area produced by the model; since the area burned in LPJ-WM is
30 nearly the same area every year, the IAV for a region is therefore forced to be small.
31 The reason why simulated fires do not appear in clusters is because the location of
32 each one is decided during assimilation based on random allocation of its point of
33 ignition and the FRI of the region; even though the FRI produced by a DVM for each

1 grid-cell depends on local climate conditions, it does so on long time scales and is
2 relatively insensitive to inter-annual variations which, for example, could cause
3 multiple fires to ignite in close proximity.

4 Refining the algorithm so that it simulates fire activity in accordance with the IAV
5 and clustering exhibited by GFED4 is a daunting task, especially as lightning, which
6 is not considered in most DVMs, is the main ignition source at these latitudes (Stocks
7 et al., 2002) and is projected to increase in frequency (Romps et al., 2014).
8 Furthermore, even though it is desirable, it is not necessary for a DVM to capture the
9 IAV and spatial variability in annual fire locations in order to make medium to long
10 term predictions on the effects of fire activity on net carbon and water fluxes. As long
11 as the FRI produced by the DVM has the correct magnitude and captures the trend in
12 fire activity in accordance with climate change, and fire size is linked to a complete
13 suite of post-fire processes, then the model is capable of accounting for the effects of
14 fire activity on an ecosystem.

15 Nevertheless, there are possible ways to modify the spatial distribution and IAV of
16 fires so they are closer to what is seen in GFED4. For example, even though LPJ-WM
17 produces annual fires whose areas rarely exceed 3% of a grid-cell, it uses climate data
18 and soil/litter properties to derive the fraction of area burned; this value increases or
19 decreases in a given year, albeit slightly, depending on how favourable the conditions
20 are for fire. This annual fluctuation could be used as an indicator of the magnitude of
21 fire activity. Instead of selecting a fire from the CCL pool and randomly assigning it
22 to grid-cells that can accommodate it (Step 6, Sect. 2.2), grid-cells that experience an
23 increase in area burned during consecutive years, or in the current year and a running
24 average of the previous ones, could be prioritized to accommodate a fire against grid-
25 cells that experience a decrease. This would allow the model not only to generate fires
26 with a realistic size distribution, as in this study, but also their location would be
27 linked to regional climatic conditions, thus further improving the fire representation.
28 The accuracy of this approach depends not only on the ability of the model to identify
29 annual fire hotspots based on climate but also on the random component of fire
30 activity. Regarding IAV, as mentioned in Step 5 of the methodology, all fires are
31 normalized so the area burned in a given year of the CCL implementation matches the
32 area burned in the original model run. This normalization was performed to preserve
33 model cohesion, but the algorithm can operate with any user-defined IAV whilst

1 maintaining FRI. Unfortunately, at these latitudes we do not yet have enough
2 information to characterize statistically the temporal behaviour of burned area (either
3 its distribution or its second order correlation properties). However, the continuing
4 availability of suitable space-based sensors will progressively fill this knowledge gap.

5 Despite the limitations described above, the assimilation methodology described here
6 gives DVMs hitherto unavailable capabilities to study post-fire behaviour under the
7 large climatic changes projected to occur in the Arctic. As long as a DVM has the
8 necessary processes to simulate post-fire dynamics (e.g. canopy radiative transfer,
9 vegetation succession, permafrost-related processes and parameterization) and is
10 correctly calibrated against field data, model runs driven by climate scenarios can
11 now offer insights into the role of fire by answering questions such as: (1) Will
12 permafrost recover after a big fire when the atmospheric temperature is rising,
13 especially in regions where it is discontinuous, and what will be the effect of the
14 projected increase in precipitation? (2) How will post-fire vegetation succession be
15 affected at ecosystem boundaries under the greening effect in the Arctic? (3) How
16 will evapotranspiration be affected under increases in fire activity and precipitation?
17 (4) How will the magnitude of fire emissions vary over sub-regions, and can changes
18 in fire activity change the sign of the land-atmosphere net carbon exchange?

19 **Acknowledgements**

20 This study was supported by the EU FP7 project EURUCAS (grant no. 295068),
21 European-Russian Centre for cooperation in the Arctic and sub-Arctic environmental
22 and climate research. E.P Kantzas acknowledges the support of Nansen Centre, St
23 Petersburg, Russia.

24 **References**

- 25 Amiro, B. D., Logan, K. A., Wotton, B. M., Flannigan, M. D., Todd, J. B., Stocks, B. J., and
26 Martell, D. L.: Fire weather index system components for large fires in the Canadian boreal
27 forest, *Int J Wildland Fire*, 13, 391-400, 2004.
- 28 Amiro, B. D., MacPherson, J. I., Desjardins, R. L., Chen, J. M., and Liu, J.: Post-fire carbon
29 dioxide fluxes in the western Canadian boreal forest: evidence from towers, aircraft and
30 remote sensing, *Agr Forest Meteorol*, 115, 91-107, 2003.
- 31 Amiro, B. D., Orchansky, A. L., Barr, A. G., Black, T. A., Chambers, S. D., Chapin, F. S.,
32 Goulden, M. L., Litvak, M., Liu, H. P., McCaughey, J. H., McMillan, A., and Randerson, J. T.:
33 The effect of post-fire stand age on the boreal forest energy balance, *Agr Forest Meteorol*,
34 140, 41-50, 2006.

1 Amiro, B. D., Stocks, B. J., Alexander, M. E., Flannigan, M. D., and Wotton, B. M.: Fire, climate
2 change, carbon and fuel management in the Canadian boreal forest, *Int J Wildland Fire*, 10,
3 405-413, 2001a.

4 Amiro, B. D., Todd, J. B., Wotton, B. M., Logan, K. A., Flannigan, M. D., Stocks, B. J., Mason, J.
5 A., Martell, D. L., and Hirsch, K. G.: Direct carbon emissions from Canadian forest fires, 1959-
6 1999, *Canadian Journal of Forest Research*, 31, 512-525, 2001b.

7 Arino, O., Bicheron, P., Achard, F., Latham, J., Witt, R., and Weber, J. L.: GLOBCOVER The
8 most detailed portrait of Earth, *Esa Bull-Eur Space*, 24-31, 2008.

9 Arino, O., Casadio, S., and Serpe, D.: Global night-time fire season timing and fire count
10 trends using the ATSR instrument series, *Remote Sens Environ*, 116, 226-238, 2012.

11 Balshi, M. S., McGuire, A. D., Duffy, P., Flannigan, M., Kicklighter, D. W., and Melillo, J.:
12 Vulnerability of carbon storage in North American boreal forests to wildfires during the 21st
13 century, *Global Change Biol*, 15, 1491-1510, 2009a.

14 Balshi, M. S., McGuire, A. D., Duffy, P., Flannigan, M., Walsh, J., and Melillo, J.: Assessing the
15 response of area burned to changing climate in western boreal North America using a
16 Multivariate Adaptive Regression Splines (MARS) approach, *Global Change Biol*, 15, 578-600,
17 2009b.

18 Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J.
19 M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O.,
20 Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator
21 (JULES), model description - Part 1: Energy and water fluxes, *Geosci Model Dev*, 4, 677-699,
22 2011.

23 Bond-Lamberty, B., Peckham, S. D., Gower, S. T., and Ewers, B. E.: Effects of fire on regional
24 evapotranspiration in the central Canadian boreal forest, *Global Change Biol*, 15, 1242-1254,
25 2009.

26 Castellanos, P., Boersma, K. F., and van der Werf, G. R.: Satellite observations indicate
27 substantial spatiotemporal variability in biomass burning NO_x emission factors for South
28 America, *Atmos Chem Phys*, 14, 3929-3943, 2014.

29 Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and
30 Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES
31 land-surface model, *Geosci. Model Dev.*, 8, 1493-1508, 2015.

32 Clark, K. L., Skowronski, N., Gallagher, M., Renninger, H., and Schafer, K.: Effects of invasive
33 insects and fire on forest energy exchange and evapotranspiration in the New Jersey
34 pinelands, *Agr Forest Meteorol*, 166, 50-61, 2012.

35 Dore, S., Kolb, T. E., Montes-Helu, M., Eckert, S. E., Sullivan, B. W., Hungate, B. A., Kaye, J. P.,
36 Hart, S. C., Koch, G. W., and Finkral, A.: Carbon and water fluxes from ponderosa pine forests
37 disturbed by wildfire and thinning, *Ecol Appl*, 20, 663-683, 2010.

38 Dore, S., Montes-Helu, M., Hart, S. C., Hungate, B. A., Koch, G. W., Moon, J. B., Finkral, A. J.,
39 and Kolb, T. E.: Recovery of ponderosa pine ecosystem carbon and water fluxes from
40 thinning and stand-replacing fire, *Global Change Biol*, 18, 3171-3185, 2012.

41 Epstein, H. E., Reynolds, M. K., Walker, D. A., Bhatt, U. S., Tucker, C. J., and Pinzon, J. E.:
42 Dynamics of aboveground phytomass of the circumpolar Arctic tundra during the past three
43 decades, *Environ Res Lett*, 7, 2012.

44 Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O.: An active-fire based burned
45 area mapping algorithm for the MODIS sensor, *Remote Sens Environ*, 113, 408-420, 2009.

46 Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual
47 burned area using the fourth-generation global fire emissions database (GFED4), *Journal of
48 Geophysical Research: Biogeosciences*, 118, 317-328, 2013.

49 Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C.,
50 and DeFries, R. S.: Assessing variability and long-term trends in burned area by merging
51 multiple satellite fire products, *Biogeosciences*, 7, 1171-1186, 2010.

1 Gonzalez, R. C., Woods, R. E., and Eddins, S. L.: Digital Image Processing Using MATLAB,
2 Prentice-Hall, Inc., 2003.

3 Harden, J. W., Manies, K. L., Turetsky, M. R., and Neff, J. C.: Effects of wildfire and
4 permafrost on soil organic matter and soil climate in interior Alaska, *Global Change Biol*, 12,
5 2391-2403, 2006.

6 Harden, J. W., Trumbore, S. E., Stocks, B. J., Hirsch, A., Gower, S. T., O'Neill, K. P., and
7 Kasischke, E. S.: The role of fire in the boreal carbon budget, *Global Change Biol*, 6, 174-184,
8 2000.

9 Jia, G. S. J., Epstein, H. E., and Walker, D. A.: Greening of arctic Alaska, 1981-2001, *Geophys*
10 *Res Lett*, 30, 2003.

11 Jiang, Y., Zhuang, Q., and Mandallaz, D.: Modeling Large Fire Frequency and Burned Area in
12 Canadian Terrestrial Ecosystems with Poisson Models, *Environmental Modeling &*
13 *Assessment*, 2012.

14 Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J. J.,
15 Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions
16 estimated with a global fire assimilation system based on observed fire radiative power,
17 *Biogeosciences*, 9, 527-554, 2012.

18 Kantzas, E., Lomas, M., and Quegan, S.: Fire at high latitudes: Data-model comparisons and
19 their consequences, *Global Biogeochem Cy*, 27, 677-691, 2013.

20 Keith, H., Mackey, B. G., and Lindenmayer, D. B.: Re-evaluation of forest biomass carbon
21 stocks and lessons from the world's most carbon-dense forests, *P Natl Acad Sci USA*, 106,
22 11635-11640, 2009.

23 Kloster, S., Mahowald, N. M., Randerson, J. T., Thornton, P. E., Hoffman, F. M., Levis, S.,
24 Lawrence, P. J., Feddema, J. J., Oleson, K. W., and Lawrence, D. M.: Fire dynamics during the
25 20th century simulated by the Community Land Model, *Biogeosciences*, 7, 1877-1902, 2010.

26 Koltunov, A., Ustin, S. L., and Prins, E. M.: On timeliness and accuracy of wildfire detection by
27 the GOES WF-ABBA algorithm over California during the 2006 fire season, *Remote Sens*
28 *Environ*, 127, 194-209, 2012.

29 Koren, I., Kaufman, Y. J., Remer, L. A., and Martins, J. V.: Measurement of the effect of
30 Amazon smoke on inhibition of cloud formation, *Science*, 303, 1342-1345, 2004.

31 Krawchuk, M. A., Moritz, M. A., Parisien, M. A., Van Dorn, J., and Hayhoe, K.: Global
32 Pyrogeography: the Current and Future Distribution of Wildfire, *Plos One*, 4, 2009.

33 Krezek-Hanes, C. C., Ahern, F., Cantin, A., and Flannigan, M. D.: Trends in large fires in
34 Canada, 1959-2007, Canadian Councils of Resource Ministers, Ottawa, 2011.

35 Li, F., Bond-Lamberty, B., and Levis, S.: Quantifying the role of fire in the Earth system - Part
36 2: Impact on the net carbon balance of global terrestrial ecosystems for the 20th century,
37 *Biogeosciences*, 11, 1345-1360, 2014.

38 Mack, M. C., Bret-Harte, M. S., Hollingsworth, T. N., Jandt, R. R., Schuur, E. A. G., Shaver, G.
39 R., and Verbyla, D. L.: Carbon loss from an unprecedented Arctic tundra wildfire, *Nature*, 475,
40 489-492, 2011.

41 Matson, M. and Dozier, J.: Identification of Subresolution High-Temperature Sources Using a
42 Thermal Ir Sensor, *Photogramm Eng Rem S*, 47, 1311-1318, 1981.

43 Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly
44 climate observations and associated high-resolution grids, *Int J Climatol*, 25, 693-712, 2005.

45 Morissette, J. T., Giglio, L., Csiszar, I., Setzer, A., Schroeder, W., Morton, D., and Justice, C. O.:
46 Validation of MODIS Active Fire Detection Products Derived from Two Algorithms, *Earth*
47 *Interact*, 9, 1-25, 2005.

48 Ping, C. L., Michaelson, G. J., Jorgenson, M. T., Kimble, J. M., Epstein, H., Romanovsky, V. E.,
49 and Walker, D. A.: High stocks of soil organic carbon in the North American Arctic region, *Nat*
50 *Geosci*, 1, 615-619, 2008.

1 Podur, J. J., Martell, D. L., and Stanford, D.: A compound Poisson model for the annual area
2 burned by forest fires in the province of Ontario, *Environmetrics*, 21, 457-469, 2010.

3 Prentice, I. C., Kelley, D. I., Foster, P. N., Friedlingstein, P., Harrison, S. P., and Bartlein, P. J.:
4 Modeling fire and the terrestrial carbon balance, *Global Biogeochem Cy*, 25, 2011.

5 Romps, D. M., Seeley, J. T., Volaro, D., and Molinari, J.: Projected increase in lightning strikes
6 in the United States due to global warming, *Science*, 346, 851-854, 2014.

7 Roy, D. P., Boschetti, L., Justice, C. O., and Ju, J.: The collection 5 MODIS burned area product
8 - Global evaluation by comparison with the MODIS active fire product, *Remote Sens Environ*,
9 112, 3690-3707, 2008.

10 Sassen, K. and Khvorostyanov, V. I.: Cloud effects from boreal forest fire smoke: evidence for
11 ice nucleation from polarization lidar data and cloud model simulations, *Environ Res Lett*, 3,
12 2008.

13 Schepaschenko, D. G., Mukhortova, L. V., Shvidenko, A. Z., and Vedrova, E. F.: The Pool of
14 Organic Carbon in the Soils of Russia, *Eurasian Soil Sci+*, 46, 107-116, 2013.

15 Sitch, S., Smith, B., Prentice, I. C., Arneeth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S.,
16 Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics,
17 plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model,
18 *Global Change Biol*, 9, 161-185, 2003.

19 Stocks, B. J., Fosberg, M. A., Lynham, T. J., Mearns, L., Wotton, B. M., Yang, Q., Jin, J. Z.,
20 Lawrence, K., Hartley, G. R., Mason, J. A., and McKenney, D. W.: Climate change and forest
21 fire potential in Russian and Canadian boreal forests, *Climatic Change*, 38, 1-13, 1998.

22 Stocks, B. J., Mason, J. A., Todd, J. B., Bosch, E. M., Wotton, B. M., Amiro, B. D., Flannigan, M.
23 D., Hirsch, K. G., Logan, K. A., Martell, D. L., and Skinner, W. R.: Large forest fires in Canada,
24 1959-1997, *J Geophys Res-Atmos*, 108, 2002.

25 Sturm, M., Racine, C., and Tape, K.: Climate change - Increasing shrub abundance in the
26 Arctic, *Nature*, 411, 546-547, 2001.

27 Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil
28 organic carbon pools in the northern circumpolar permafrost region, *Global Biogeochem Cy*,
29 23, 2009.

30 Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C.:
31 The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas
32 emissions: results from a process-based model, *Biogeosciences*, 7, 1991-2011, 2010.

33 Thonicke, K., Venevsky, S., Sitch, S., and Cramer, W.: The role of fire disturbance for global
34 vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model, *Global Ecol*
35 *Biogeogr*, 10, 661-677, 2001.

36 Valentini, R., Arneeth, A., Bombelli, A., Castaldi, S., Gatti, R. C., Chevallier, F., Ciais, P., Grieco,
37 E., Hartmann, J., Henry, M., Houghton, R. A., Jung, M., Kutsch, W. L., Malhi, Y., Mayorga, E.,
38 Merbold, L., Murray-Tortarolo, G., Papale, D., Peylin, P., Poulter, B., Raymond, P. A., Santini,
39 M., Sitch, S., Laurin, G. V., van der Werf, G. R., Williams, C. A., and Scholes, R. J.: A full
40 greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities,
41 *Biogeosciences*, 11, 381-407, 2014.

42 van der Werf, G. R., Randerson, J. T., Collatz, G. J., Giglio, L., Kasibhatla, P. S., Arellano, A. F.,
43 Olsen, S. C., and Kasischke, E. S.: Continental-Scale Partitioning of Fire Emissions During the
44 1997 to 2001 El Niño/La Niña Period, *Science*, 303, 73-76, 2004.

45 van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton,
46 D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution
47 of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), *Atmos Chem Phys*,
48 10, 11707-11735, 2010.

49 Viereck, L. A.: The effects of fire in black spruce ecosystems of Alaska and northern Canada.
50 In: *The role of fire in northern circumpolar ecosystems*, Wien, R. W. and MacLean, D. A.
51 (Eds.), Wiley & Sons Ltd., 1983.

1 Viereck, L. A. and Dyrness, C. T.: Ecological effects of the Wickersham Dome fire near
2 Fairbanks, Alaska, US For. Serv.PNW-90, 71 pp., 1979.

3 Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic
4 global vegetation model: 1. Evaluation and sensitivity of physical land surface processes,
5 Global Biogeochem Cy, 23, 2009a.

6 Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic
7 global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle
8 processes, Global Biogeochem Cy, 23, 2009b.

9 Wiitala, M. R.: Assessing the risk of cumulative burned acreage using the Poisson probability
10 model, Gen. Tech. Rep. PSW-GTR-173, 1999.

11 Wooster, M. J., Xu, W., and Nightingale, T.: Sentinel-3 SLSTR active fire detection and FRP
12 product: Pre-launch algorithm development and performance evaluation using MODIS and
13 ASTER datasets, Remote Sens Environ, 120, 236-254, 2012.

14 Wooster, M. J. and Zhang, Y. H.: Boreal forest fires burn less intensely in Russia than in North
15 America, Geophys Res Lett, 31, 2004.

16 Xu, L., Myneni, R. B., Chapin, F. S., Callaghan, T. V., Pinzon, J. E., Tucker, C. J., Zhu, Z., Bi, J.,
17 Ciais, P., Tommervik, H., Euskirchen, E. S., Forbes, B. C., Piao, S. L., Anderson, B. T., Ganguly,
18 S., Nemani, R. R., Goetz, S. J., Beck, P. S. A., Bunn, A. G., Cao, C., and Stroeve, J. C.:
19 Temperature and vegetation seasonality diminishment over northern lands, Nature Climate
20 Change, 3, 581-586, 2013.

21 Yoshikawa, K., Bolton, W. R., Romanovsky, V. E., Fukuda, M., and Hinzman, L. D.: Impacts of
22 wildfire on the permafrost in the boreal forests of Interior Alaska, J Geophys Res-Atmos, 108,
23 2002.

24

25

26

27

28

29

30

31

32

33

34

35

36

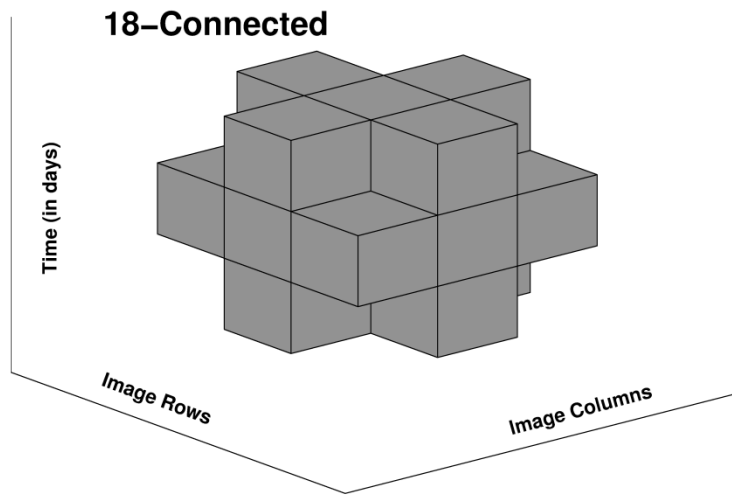
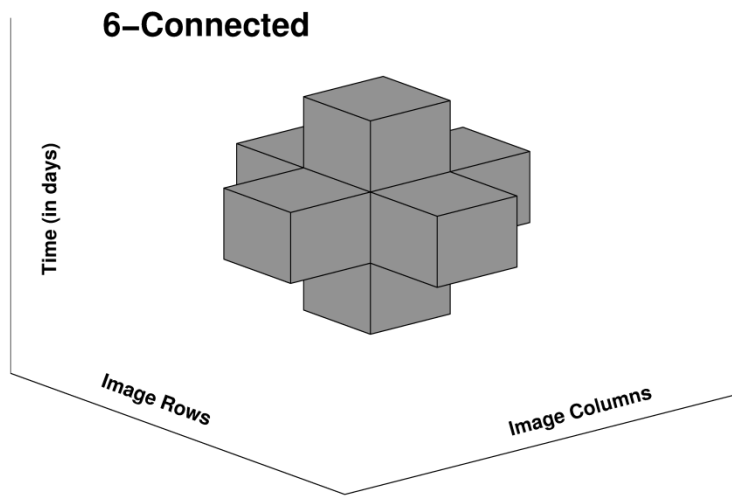
37

38

39

1 **Figures**

2

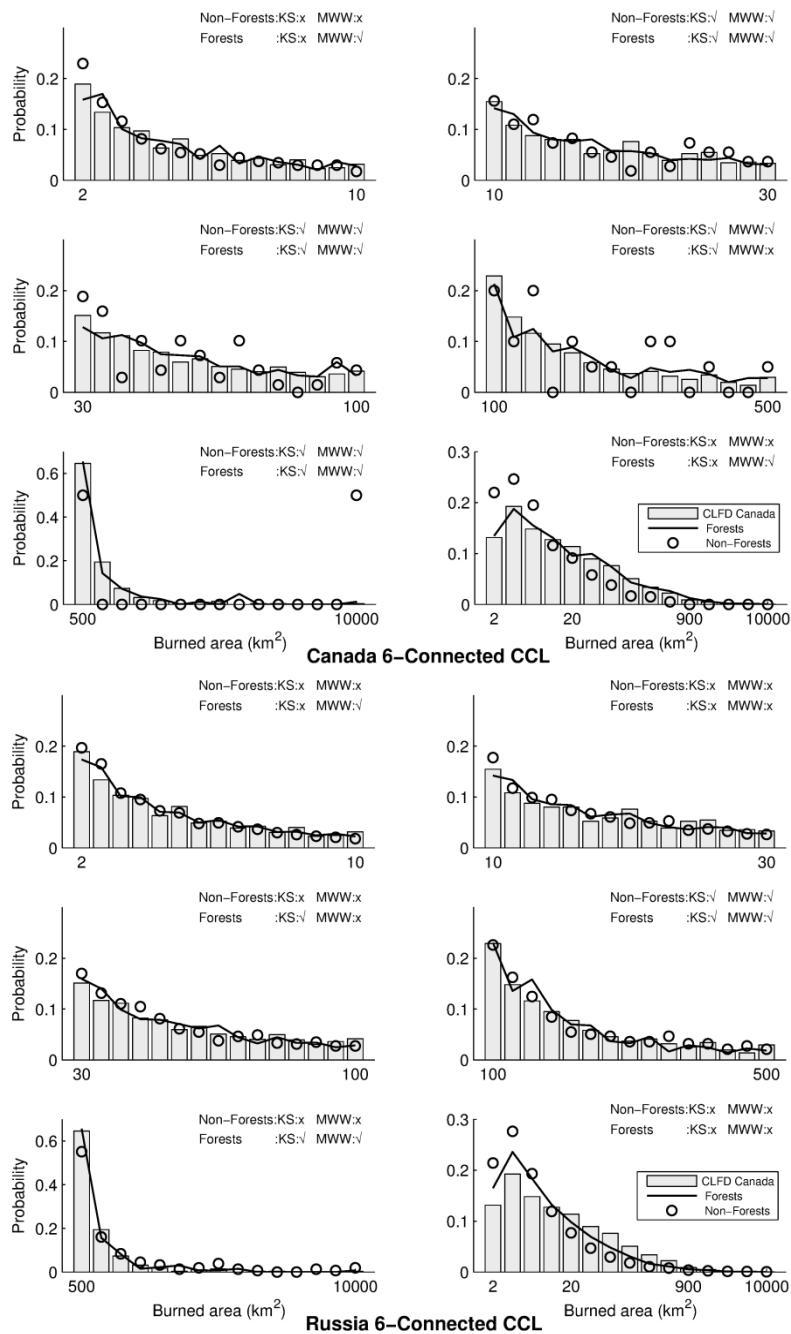


3

4

5 Figure 1: 6- and 18-connected pixel connectivity in 3-dimensional CCL analysis; the
6 axes are image row, image column and day of the year.

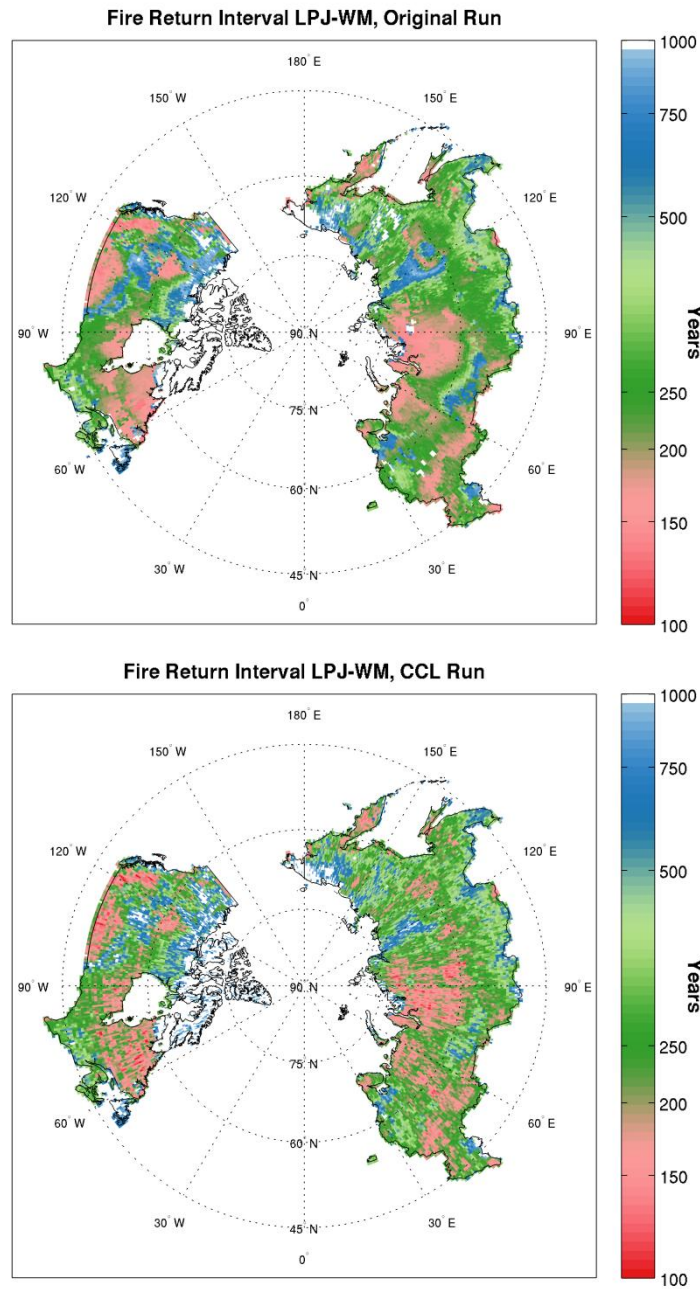
7



1

2 Figure 2: Histograms of area burned in each fire size category obtained from the
 3 CLFD and the application of 6-connected CCL to the GFED burned area daily
 4 product. The CLFD results only describe forest fires in Canada, while the CCL-6
 5 results are given separately for forests and non-forests in Canada (top) and Russia
 6 (bottom). The limits of the x-axes in each figure give the range of burned area studied
 7 in each category; the x-axis in the bottom right figure for each region uses a
 8 logarithmic scale. A tick or cross shows whether the CCL-derived distributions for
 9 forests and non-forests pass or fail the Kolmogorov-Smirnov (KS) and Mann-
 10 Whitney-Wilcoxon (MWW) test when compared to their size-respective CLFD
 11 distributions.

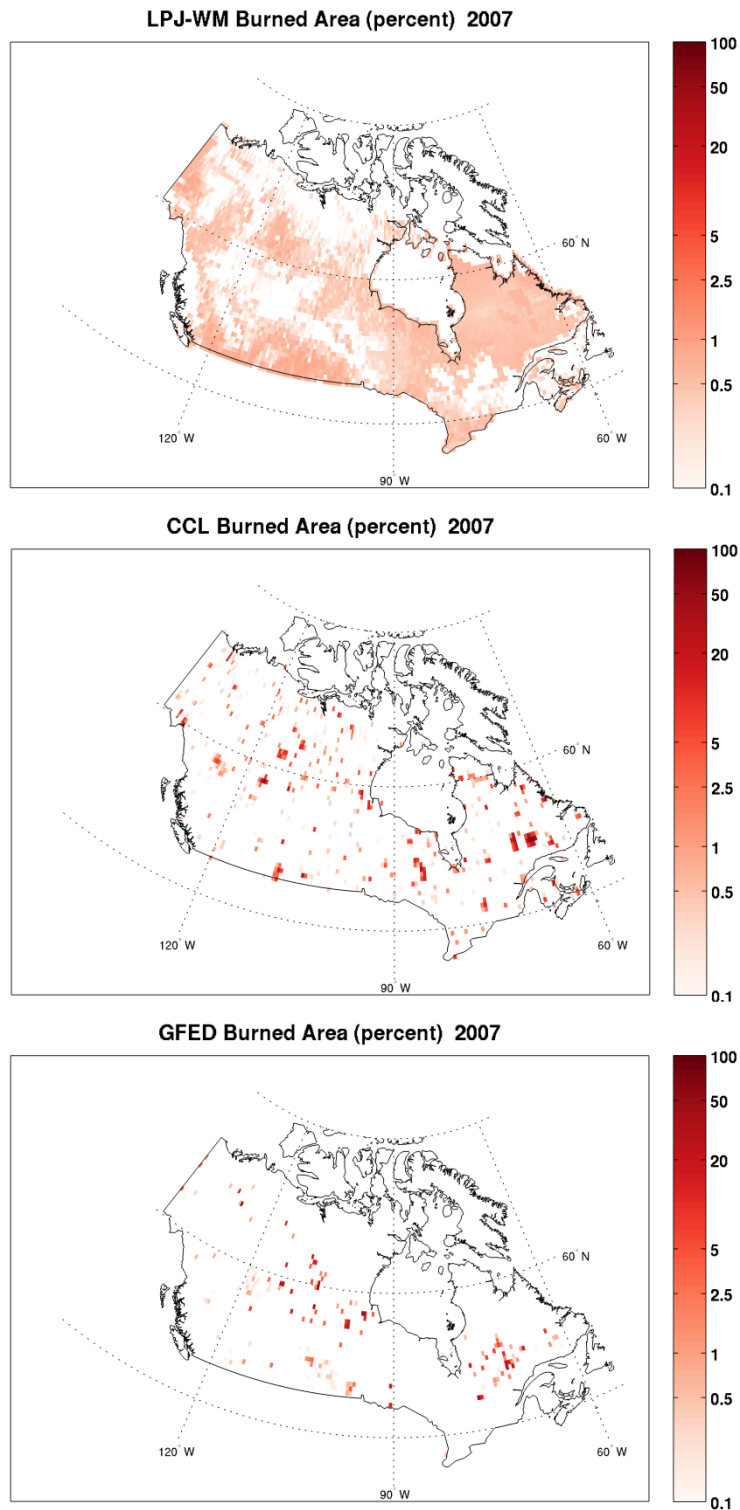
1



2

3 Figure 3: (top) Fire Return Interval produced by an original LPJ-WM run over a 1000
4 years spin-up combined with a transient run (1901-2012) for Canada and Russia.
5 (bottom) FRI produced by a LPJ-WM run over the same period with the CCL
6 methodology.

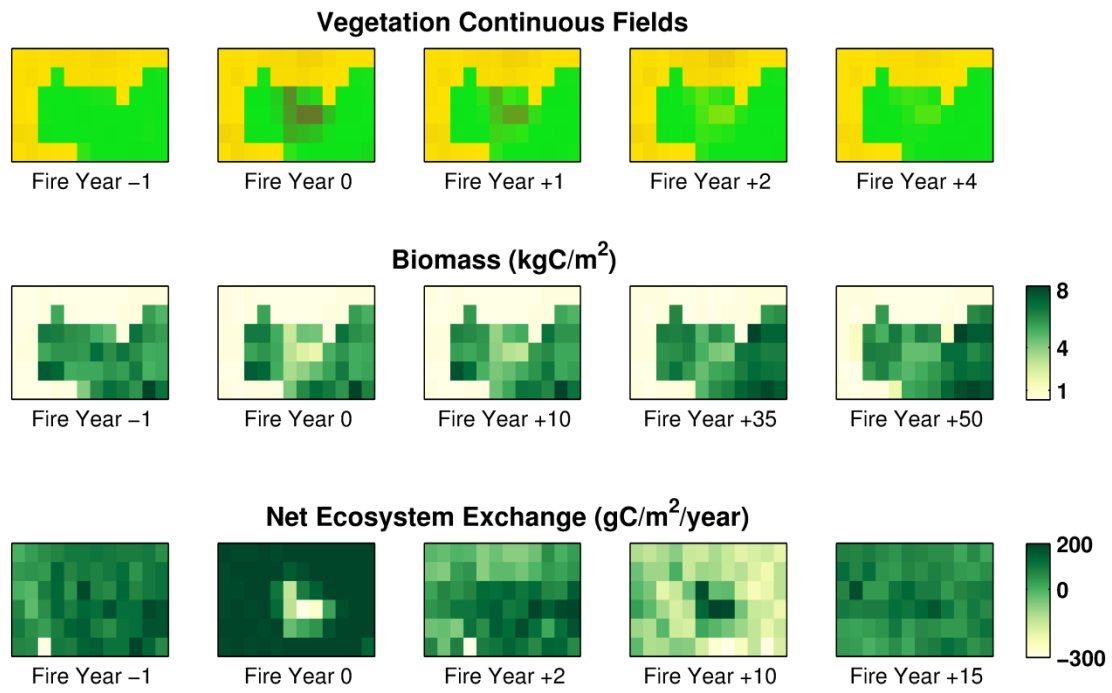
7



1

2 Figure 4: Fractional burned area per grid-cell (%) for 2007 in an original LPJ-WM run
 3 (top), a CCL run (centre) and GFED (bottom). Note that, since the fire is stochastic in
 4 the CCL run, different runs would produce different fires for the same year but the
 5 overall fraction of area burned would remain equal to that of the original run.

1



2

3 Figure 5: Post-fire evolution of the carbon stocks and fluxes after the fire disturbance
4 indicated in Fig. 4 (centre). Top: vegetation cover in Vegetation Continuous Fields
5 format (Green = Trees, Yellow = Grass, Brown = Bare Ground); middle: biomass
6 density in kg of carbon m⁻²; bottom: Net Ecosystem Exchange in g of carbon m⁻² y⁻¹.

7

8