Using Endorsement Theory to Validate Annual Burnt Areas of a Large Region in Central Siberia

Charles George[1], R. Wadsworth[1], Y. Flety[2], A. Thomson[1], C. Rowland[1],
F. Gerard[1], H. Balzter[1]

[1]Centre for Ecology and Hydrology (CEH) Monks Wood, Abbots Ripton, Huntingdon, Cambridgeshire, PE28 2LS, UK
[2]Dept of Geography, University of Lyon, Le bois de Thuellin 38630 Veyrins-Thuellin France.

email: ctg@ceh.ac.uk

Abstract – Validation of remote sensing products can be very difficult, especially if the study area is large and inaccessible. In this paper, we move away from the more traditional methods of validation, to investigate whether a non-numerical approach can be utilized. Endorsement theory is useful in this context because it allows inference to be made from partial (and incomplete) knowledge, and can represent common knowledge in a natural form without being necessary to translate an experts' opinion into a numerical value. In this study we use endorsement theory to validate a fire disturbance map covering over 3 million km^2 and 12 years using four land cover maps from years 2000 to 2004. The results show that while the method shows potential, as with other validation techniques, the results are dependent on the quality of the reference data.

Keywords: Validation, Endorsement theory, Burnt areas, Land cover.

1. INTRODUCTION

The Russian forests make up 22 percent of the world forests, and contain up to 65% of the worlds closed boreal forest, and have, according to some authorities (Snyder, 2004) a strong influence on the global climate. Fire disturbances in particular contribute significantly to the transfer of large amounts of carbon to the atmosphere and ecosystem deposits. According to official statistics, between 18,000 and 37,000 fires, affecting 0.48 to 1.2 million ha of forest in Russia from 1950-1999 (Shvidenko 2001), with increasing areas being burnt in more recent years.

In order to improve our understanding of boreal forest fires and its impact on the global biogeochemical cycles, we need spatial data describing fire dynamics in terms of frequency and size. Fire statistics for the past 20 years are available for most northern hemisphere countries, but for the Siberian Boreal forest official fire statistics are incomplete and biased. A forest disturbance map was produced within the framework of the EC - GMES funded SIBERIA-II project which required quantitative measures of the geographic extent and temporal distribution of forest fire disturbances as far back in time as possible, for different forest types in Central Siberia. The data is to be incorporated into the International Institute for Applied Systems Analysis (IIASA) green house gas accounting system (Nilsson, 2003) with the aim of improving estimates of carbon and nitrogen based greenhouse gas emissions from forest fires. For the fire disturbance map to have any real value, however, a validation needs to be carried out. The validation of this product is difficult using traditional methods. Ground data is in very short supply due to the vastness of the area, and minimal communication routes. An additional problem is the required spatial resolution; the disturbance map is at 1km resolution and any comparison with high resolution imagery, such as Landsat (E)TM, is limited due to the temporal distribution of the fire scars ranging from 1992 to 2003, and the lack of availability of suitable cloud free imagery. Comparisons with other burnt area products such as GBA2000 or GLOBSCAR are not necessarily appropriate, as they are only for one year, and are subject to the same validation problems.

Another approach is Endorsement Theory which is a non-numerical approach to combining incomplete and uncertain evidence, where each extra additional pieces of information, endorses a step in the argument if it satisfies certain requirements (Cohen 1985).

Endorsement theory is useful in this context because:

a) It allows inference to be made from partial (and incomplete) knowledge.
b) The reasoning process is explicit and highly heuristic.
c) It can represent common knowledge in a natural form and it is not necessary to translate an experts' opinion into a numerical value.

The basic premise we are adopting is, if after what we have identified as a fire disturbance, there is a land cover change from forest to, for example grassland or cropland, then this is supporting evidence for that area having burnt. However if the land cover doesn't change, or changes to something such as wetland, then there is disbelief that this is a fire disturbance.

2. METHODS

Study area

The full Siberian study area is 3 million km² between 52^o-72^o Northern latitude and 88^o-110^o Eastern longitude (Fig. 1) and includes all the bio-climatic zones and major land classes of Northern Eurasia. According to the IGBP DISCover land cover classification version 2, derived from the Global Land Cover Characteristics Database approximately 50% of the area is forest, 32 % is woody savanna or shrublands, 7% is barren or sparse vegetation. The rest of the area consists of grasslands, croplands, wetlands, urban and water. Only those areas that were classified by the IGBP land cover map as forest, woody savanna or shrublands were included in the disturbance classification.

Fire Disturbance Map

Two different procedures used MODIS/Terra Nadir BRDF-Adjusted Reflectance 16-Day composites (MOD43B4) acquired for the summers of 2001, 2002 and 2003, to identify burnt areas which were created either in the same year of image acquisition (algorithm 1) or areas which burnt up to 10 years prior to the image acquisition date (algorithm 2). The former being an Normalized Difference Vegetation Index (NDVI) differencing approach , the latter involved using the NDSWIR (Normalized Difference ShortWave InfraRed index), which is sensitive to canopy moisture content and therefore canopy density (Gerard 2003). The two algorithms were needed as the composite image dates for the NDSWIR method were distributed across the fire season of 2003, so fires occurring later in 2003 may not have been identified. Consequently the NDVI-derived fire scar areas for 2003 were added to the NDSWIR derived burned area map. In both cases the IGBP land cover map derived from 1992 and 1993 AVHRR imagery was used to exclude the non-woody areas, and hotspot (thermal anomalies) information was used to date the burnt areas.

Figure 1. SIBERIA-2 study area (highlighted).

Land Cover Maps

The land cover maps (LCMs) used imagery from several different dates. The IGBP DISCover map is derived from 1km Advanced Very High Resolution Radiometer (AVHRR) data spanning a 12-month period (April 1992-March 1993). It was developed using monthly NDVI composites in an unsupervised classification. The GLC2000 LCM used daily SPOT-VEGETATION imagery from November 1999 to December 2000, and used the ISODATA clustering algorithm, with subsequent amendment by specialized vegetation indices to resolve the classes. For the years 2001 to 2004 LCMs were produced as part of the SIBERIA-2 project. MODIS 8-day temporal composites were independently classified using imagery from May to October within a C5.0 decision tree classifier. The independent thematic results were then combined using a Bayesian addition to create a single land-cover map for each year. Even though the same methodology was used for each of the SIBERIA maps, each map could be said to be independent, as the results from one year had no effect on the results for the other years.

Endorsement Process

In the semantic-statistical approach developed by Comber *et al* (2004a,b) to explore the consistency between heterogeneous land-cover data, the scores are converted to proportions, and then combined with Dempster-Shafer values/statistics. In our case, with five data sets (rather than two) and a single issue (has it burnt or not) a simpler rule was applied.

Only the polygons identified as having burnt before 2000 were processed as any that were burnt after this time would not show up as a land cover change in all of the LCMs. For each of the fire disturbance polygons a cross tabulation was carried out against each of the LCMs to count the pixel numbers of each land cover type per polygon. The land cover types for each year were then aggregated depending on whether it was thought that they could be representing a possible biomass decrease caused by fire. The three codes used were :

1 Belief that the land cover change could have been caused by fire.
0 Uncertain or no evidence for a fire, and
-1 All the evidence suggests no fire could have occurred.

Any assignment is made on the basis that the land cover maps are 100% correct, and that the effect of the fire will be detectable the following year. Note only some misclassifications will affect the result, for example misclassification within any of the nine GLC2000 forest classes has no effect on the result The assigned belief levels are given in Table 1.

Table 1. The Codes Given to Each Land Cover

GLC 2000	Score	UWS 2001/2/3/4	Score
1-9 Forest	-1	1 Water	-1
10 Humid Grassland	1	2 Barren Ground	-1
11 Steppe	1	3 Urban	1
12 Bogs Marsh	0	4 Croplands	1
13 Palsa Bogs	0	5 Cropland/Forest Complex	1
14 Riparian Vegetation	1	6 Evergreen Needleleaf	-1
15 Barren Tundra	1	7 Deciduous Broadleaf	-1
16 Prostrate Shrub Tundra	1	8 Needleleaf/Broadleaf Forest	-1
17 Sedge Tundra	1	9 Mixed Forest	-1
18 Shrub Tundra	1	10 Broadleaf/Needleleaf Forest	-1
19 Recent Burns	1	11 Deciduous Needleaf Forest	-1
20 Cropland	1	12 Humid Grassland	1
21Forest Nat Veg Cplx	1	13 Wetland	0
22 Forest Cropland Cplx	1	14 Steppe	1
23 Cropland Grassld Cplx	1	15 Tundra Lichen-Moss	1
24 Bare soil Rock	-1	16 Tundra Heath	1
25 Permanent Snow Ice	-1		
26 Water	-1		
27 Urban	1		
28 Tundra / Shrub?	1		
29 Salt Pans	-1		

The percentages by pixel of each belief code within a polygon were then calculated. Those polygons with 60% or more represented by a single code were then assigned that code. This process resulted in the individual polygons having 1, 0 or -1 for each of the land cover maps.

The interaction of these beliefs produces the overall endorsements. These are

- **Definite** – evidence provides conclusive belief (4-5 positive belief values) and no disbelief
- **Confident** – evidence provides prima-facie belief (3 positive belief values) and no *prima-facie* disbelief

- **Likely** – strong belief > strong disbelief (2 positive belief values)
- **Indicated** - weak belief > weak disbelief (1 positive belief value)
- **Contradicted** – evidence provides conclusive disbelief. (>=4 negative belief values, including at least 1 from the GLC).

3. RESULTS

Out of the 1156 polygons that were burnt before 2000, 21 were not scored by the Endorsement theory as they did not contain a dominant land cover. Of the remaining 1135, 518 (46%) were definitive and 496 (44%) contradicted (Table 1). If the confident and likely results are added to the definitive, almost 54% of fire disturbances are 'endorsed' by this process.

Table 2. Results from the Endorsement Process

	definitive	confident	likely	indicated	contradicted
	evidence occurs 4-5 times	evidence occurs 3 times	evidence occurs twice	evidence occurs once	no evidence
Numbers of polygons	518	52	40	29	496
% of polygons	45.6	4.6	3.5	2.6	43.7
aggregated %	53.74			2.6	43.7

4. DISCUSSION

At first glance, the results suggest that only half of the fire disturbances identified within the SIBERIA-2 project are endorsed by the land cover maps. However all the polygons examined had active fires present at some time between 1992 and 1999, and even if a few cases were industrial, agricultural or logging fires, the great majority would have been forest fires. Any inconsistencies may have been caused by the two assumptions underlying this method. The first, that the land cover maps are 100% correct, and the second, that the effect of the fire would be noticeable soon after the event (there would be no further vegetation mortality in the years following the fire). The consequence of the first assumption would have the greater significance. To illustrate, there are many polygons that have been flagged as Contradicted even though according to other sources they are fire disturbances. The most compelling of these other sources are the thermal anomalies, especially if they are clustered, or several appear in close proximity, reducing the possibility that they may be false positives. For example even though polygons B and C (figure 2) contain many thermal anomalies and the GLC2000 map shows a few recently burnt pixels they are both contradicted, largely due to the SIBERIA-2 maps classing the polygons as Deciduous Needleleaf. This classification of the fire disturbances as *Larix* species is relatively common. Either these land cover types burn more readily, or the classifier is getting confused by the regenerating vegetation which for this age of disturbance would consist of deciduous shrubs. This could be evidence that the initial beliefs may need altering, although further investigation is needed. For the polygons classified as definitive fire disturbances (polygon A) the results seem to be reasonably conclusive, with the LCM's not only endorsing the disturbance map with their land cover types but also reinforcing the shapes of the polygons.

5. CONCLUSION

The endorsement theory of using several different data sources for validating a remote sensing product, where traditional methods are limited, has potential. However, care should be taken in choosing the endorsing data sources, as their uncertainties and limitations need to be well documented and understood. Ideally, the the allocation of the initial beliefs should be independently decided by three or more experts as there is an element of subjectivity for some of the allocations.

6. REFERENCES

Cohen, P. (1985). "Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach." Pitman Publishing Ltd, London.

Comber A., F. P., Wadsworth R. (2004a). "A semantic statistical approach to negotiation heterogeneous ontologies." *Int Jo Geographic Information Science* **18**(7): 691-708.

Comber A., F. P., Wadsworth R., (2005). "You know What Land Cover is but does Anyone Else?....An Investigation into Semantic and Ontological Confusion." INT. J. REMOTE SENSING **26**(1): 223-228.

Comber A., F. P., Wadsworth, R.A. (2004b). "Assessment of a semantic statistical approach to detecting land cover change using inconsistent data sets." Photogrammetic Engineering and Remote Sensing **70**(8): 931-938.

Comber A. J., L. A., Lishman J., (2004). "A Comparison of Bayes',Dempster-Shafer and Endorsement Theories for Managing Knowledge Uncertainty in the Context of Land Cover Monitoring." Computors, Environment and Urban Systems **28**: 311-327.

Gerard, F., S. Plummer, et al. (2003). "Forest Fire Scar Detection in the Boreal Forest with Multi-temporal SPOT_VEGETATION Data." "Ieee Transactions On Geoscience And Remote Sensing **41**(11): 2575-2585".

Nilsson, S., A. Shvidenko, et al. (2003). "Quantification of Full Terrestrial Biota Major reenhouse Gases Budget at a Regional Scale: A Combination of odeling Systems, Geographical Information Systems and Remotely Sensed Data." Proceedings of the IGARSS2003 Conference – 2003 EEE International Geoscience and Remote Sensing Symposium, 21-25 July 2003, Toulouse, France.

Shvidenko, A., Goldammer, J. G., (2001). "Fire Situation in Russia." International Forest Fire News (IFFN) **24**: 41-59.

Snyder, P. K., C. Delire, et al. (2004). "Evaluating the influence of different vegetation biomes on the global climate." Climate Dynamics **3**(4): 279-302.

Table 3. Examples of the endorsement process – the letters correspond to figure 2.

	2000		2001		2002		2003		2004		
Polygon	**Major Landcover**	**Belief**	**Major Landcover**	**Belief**	**Major Landcover**	**Belief**	**Major Landcover**	**Belief**	**Major Landcover**	**Belief**	**Result**
A	decid needle	-1	cropland/ forest complex	1	cropland/ forest complex	1	cropland/ forest complex	1	cropland/ forest complex	1	definitive
B	decid needle	-1	decid needle	-1	decid needle	-1	tundra heath/lichen	1	decid needle	-1	contradicted
C	decid needle	-1	decid needle	-1	decid needle	-1	decid needle	-1	decid needle	-1	contradicted
D	inconclusive	0	cropland/ forest complex	1	decid needle	-1	inconclusive	0	decid needle	-1	indicated
E	decid needle	-1	cropland/ forest complex	1	inconclusive	0	cropland/ forest complex	1	cropland/ forest complex	1	confident

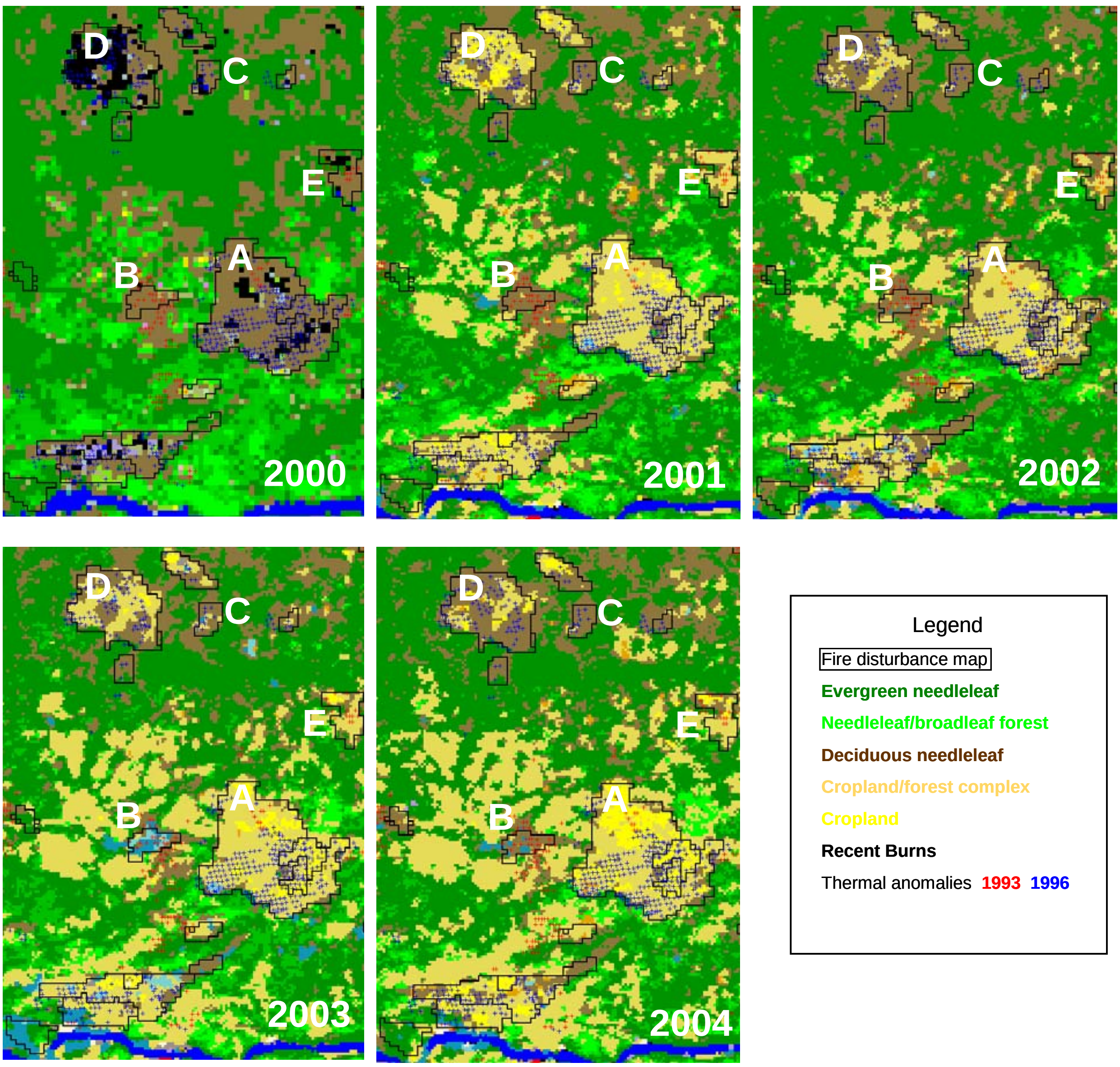

Figure 2. The five LCM's used in the Endorsement Process