Supplemental Information

to accompany

Landsat-based detection of mast events in white spruce (*Picea glauca*) forests

Matthew Garcia*, Benjamin Zuckerberg, Jalene M. LaMontagne, and Philip A. Townsend

*corresponding author: matt.e.garcia@gmail.com

Submitted to *Remote Sensing of Environment* (Elsevier) on 27 April 2020

Revision 1 submitted on 27 October 2020

Revision 2 submitted on 21 December 2020

Contents (in order of reference in the main text)

Appendix S1: Software description

Our Landsat preprocessing procedures for topographic correction, image masking, VI calculations, and data organization used *Python* v3.8.5 (Oliphant, 2007; Millman and Aivazis, 2011; Pérez et al., 2011) and several libraries including *numpy* v1.19.1 (Harris et al., 2020) and *scipy* v1.5.2 (Virtanen et al., 2020), *GDAL* v3.1.2 (Warmerdam, 2008) and *pyproj* v2.6.1 (Snow et al., 2020), and *h5py* v2.10.0 (Collette, 2013) for the HDF5 data format (HDF Group, 1997). Our procedures to query and extract VI values at the selected study sites used several additional Python libraries including *pandas* v1.1.1 (McKinney et al., 2010), *pykml* v0.2.0 (https://pythonhosted.org/pykml/), and *shapely* v1.7.1 (https://shapely.readthedocs.io/).

We developed our statistical analyses in *Python* using *Jupyter* v1.0.0 notebooks (Kluyver et al., 2016; Randles et al., 2017; Perkel, 2018; Wofford et al., 2020), a browser-based graphical extension of *IPython* v7.18.1 (Pérez and Granger, 2007). These *Jupyter*-based analyses used a number of *Python* libraries: *numpy*, *pandas*, *pingouin* v0.3.8 (Vallat, 2018), *scipy*, *scikit-learn* v0.23.2 (Pedregosa et al., 2011), and *statsmodels* v0.11.1 (Seabold and Perktold, 2010). We generated all of the figures in this work, except for the photograph in Figure 1, using *matplotlib* v3.3.1 (Hunter, 2007) and *seaborn* v0.10.1 (Waskom et al., 2020). We have made our extracted and processed Landsat VI datasets and *Python*/*Jupyter* statistical analysis notebooks openly available at https://github.com/megarcia/spruce_masting.

References

Collette, A., 2013: *Python and HDF5*. O'Reilly Media, ISBN 9781449367831.

Harris, C.R., and 24 coauthors, 2020: Array programming with NumPy. *Nature*, **585**, 357–362, doi: 10.1038/s41586-020-2649-2.

HDF Group, 1997: *Hierarchical Data Format, Version 5*. http://www.hdfgroup.org/HDF5.

Hunter, J.D., 2007: Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, **9**, 90–95, doi: 10.1109/MCSE.2007.55.

Kluyver, T., B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and the Jupyter Development Team, 2016: Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, Amsterdam: IOS Press, 87–90, doi: 10.3233/978-1-61499-649-1-87.

McKinney, W., 2010: Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, Austin, Texas, 56–61, doi: 10.25080/Majora-92bf1922-00a.

Millman, K.J., and M. Aivazis, 2011: Python for scientists and engineers. *Computing in Science Engineering*, **13**, 9–12, doi: 10.1109/MCSE.2011.36.

Oliphant, T.E., 2007: Python for scientific computing. *Computing in Science Engineering*, **9**, 10–20, doi: 10.1109/MCSE.2007.58.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, 2011: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830, https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

Pérez, F., and B.E. Granger, 2007: IPython: A system for interactive scientific computing. *Computing in Science Engineering*, **9**, 21–29, doi: 10.1109/MCSE.2007.53.

Pérez, F., B.E. Granger, and J.D. Hunter, 2011: Python: An ecosystem for scientific computing. *Computing in Science Engineering*, **13**, 13–21, doi: 10.1109/MCSE.2010.119.

Perkel, J.M., 2018: Why Jupyter is data scientists' computational notebook of choice. *Nature*, **563**, 145–146, doi: 10.1038/d41586-018-07196-1.

Randles, B.M., I.V. Pasquetto, M.S. Golshan, and C.L. Borgman, 2017: Using the Jupyter notebook as a tool for open science: An empirical study. *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–2, https://ieeexplore.ieee.org/document/7991618.

Seabold, S., and J. Perktold, 2010: Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, Austin, Texas, 92–96, doi: 10.25080/Majora-92bf1922-011.

Snow, A.D., and 29 co-authors: pyproj4/pyproj 2.6.1 Release. Zenodo, doi: 10.5281/zenodo.3783866.

Vallat, R., 2018: Pingouin: Statistics in Python. *Journal of Open Source Software*, **3**, 1026, doi: 10.21105/joss.01026.

Virtanen, P., and 33 coauthors, 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**, 261–272, doi: 10.1038/s41592-019-0686-2.

Warmerdam, F., 2008: "The Geospatial Data Abstraction Library." In: Hall G.B. and Leahy M.G. (eds), *Open Source Approaches in Spatial Data Handling*. Advances in Geographic Information Science, vol 2. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-74831-1_5.

Waskom, M., and 29 co-authors, 2020: mwaskom/seaborn: v0.10.1 (April 2020) Zenodo, doi: 10.5281/zenodo.3767070

Wofford, M.F., B.M. Boscoe, C.L. Borgman, I.V. Pasquetto, and M.S. Golshan, 2020: Jupyter notebooks as discovery mechanisms for open science: Citation practices in the astronomy community. *Computing in Science Engineering*, **22**, 5–15, doi: 10.1109/MCSE.2019.2932067.

Table S1: Example *statsmodels* prediction table result from fitting a logistic regression model to the observed record of masting across site-years. This table represents the prediction result for a number of logistic models that are discussed in the text, including the null model.

		Pred.	
		0	**1**
Obs.	**0 (non-mast)**	85	0
	1 (mast)	16	0

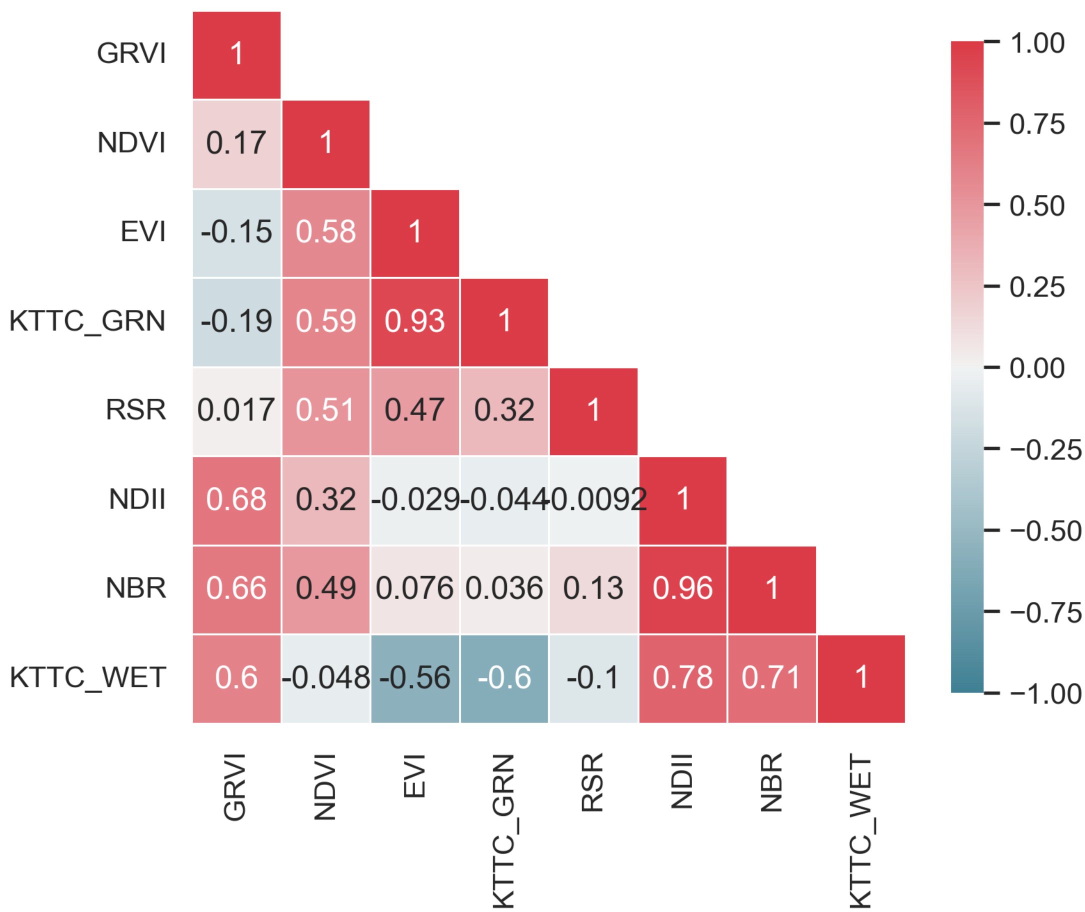

Figure S1: Correlation matrix across time series of standardized annual phenological regression slopes for all VIs used in this work.

Table S2: Results for all restricted two-VI logistic regression models using standardized annual slopes for color- and moisture-oriented VIs (see Tables 3 and 4 in the main paper). Model weights incorporate all listed models except the null model, disregarding *ΔAICc* values. See also Figures 7 and S2 for model coefficients in the top results (*ΔAICc* ≤ 2) listed here. Note that correctly predicted non-mast years are not listed but are included in the calculation of overall accuracy and Cohen's κ.

Logistic model VIs	pseudo-R^2	AICc	ΔAICc	Model weight	Correct mast years	Missed mast years	False alarms	Overall accuracy	Cohen's κ
NDVI + NBR	**0.188**	**82.1**	—	**0.378**	**4**	**12**	**0**	**0.881**	**0.359**
NDVI + NDII	0.173	83.4	+1.30	0.197	2	14	1	0.851	0.169
GRVI + NDII	0.161	84.5	+2.40	0.114	3	13	1	0.861	0.253
GRVI + NBR	0.143	86.1	+3.97	0.052	3	13	1	0.861	0.253
KTTC_GRN + NDII	0.140	86.4	+4.27	0.045	3	13	1	0.861	0.253
EVI + NDII	0.136	86.7	+4.62	0.037	3	13	1	0.861	0.253
KTTC_GRN + NBR	0.133	86.9	+4.85	0.033	3	13	1	0.861	0.253
EVI + NBR	0.130	87.2	+5.10	0.030	3	13	1	0.861	0.253
GRVI + KTTC_WET	0.128	87.4	+5.28	0.027	3	13	2	0.851	0.227
EVI + KTTC_WET	0.128	87.4	+5.31	0.027	3	13	1	0.861	0.253
NDVI + KTTC_WET	0.127	87.5	+5.39	0.025	3	13	2	0.851	0.227
KTTC_GRN + KTTC_WET	0.123	87.9	+5.77	0.021	3	13	1	0.861	0.253
GRVI + RSR	0.098	90.1	+7.97	0.007	0	16	1	0.831	-0.019
NDVI + RSR	0.072	92.3	+10.25	0.002	0	16	0	0.842	0.000
KTTC_GRN + RSR	0.072	92.4	+10.26	0.002	0	16	0	0.842	0.000
EVI + RSR	0.059	93.4	+11.36	0.001	0	16	0	0.842	0.000
null	0.011	93.7	+11.61	—	0	16	0	0.842	0.000

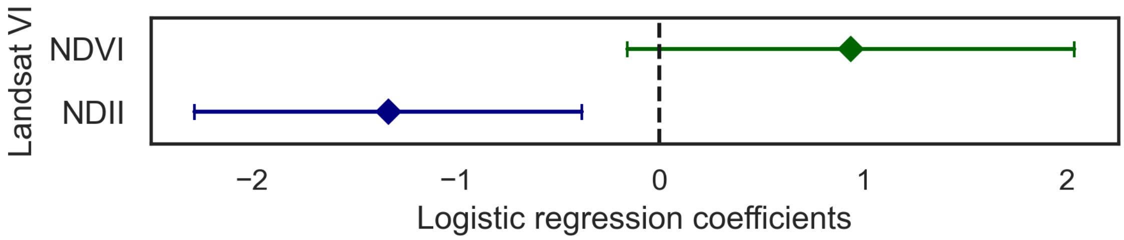

Figure S2: LMM logistic regression coefficients and 95% confidence intervals for the restricted two-VI model. See Tables 5 and S2 for model accuracy metrics.

Table S3: Averaged two-VI logistic model prediction results. Correctly predicted non-mast years (comprising 85 site-years) are not listed here. Model accuracy metrics are listed in Table 6.

Site	**Correctly predicted mast years (3)**	**Missed mast years (Type 2 error) (13)**	**False alarms (Type 1 error) (0)**
CHITTY	—	1998, 2010, 2014	—
KLOO	1993, 2010	1998, 2005, 2014	—
SILVER	—	2005, 2010, 2014	—
SULPHUR	1993	1998, 2005, 2010, 2014	—

Table S4: Results for unrestricted multi-VI logistic regression models. Only those models with $\Delta AICc < 2$ are listed. The average model and the model with the highest value of Cohen's κ are both shown in bold. Note that correctly predicted non-mast years are not listed but are included in the calculations of overall accuracy and Cohen's κ. See Figures S3 and S4 for model coefficients and Table S5 for results of the best model prediction.

Logistic model VI components	**pseudo-R^2**	**AICc**	**ΔAICc**	**Model weight**	**Correct mast years**	**Missed mast years**	**False alarms**	**Overall accuracy**	**Cohen's κ**
RSR + NBR	0.235	78.0	—	0.177	3	13	1	0.861	0.253
GRVI + RSR + NBR	**0.252**	**78.7**	**+0.707**	**0.124**	**4**	**12**	**0**	**0.881**	**0.359**
NDVI + RSR + NBR	0.251	78.8	+0.779	0.120	3	13	1	0.861	0.253
EVI + KTTC_GRN + RSR + NBR	0.270	79.4	+1.392	0.088	3	13	3	0.842	0.204
RSR + NDII	0.219	79.4	+1.396	0.088	2	14	2	0.842	0.146
EVI + KTTC_GRN + RSR + KTTC_WET	0.270	79.4	+1.416	0.087	3	13	5	0.822	0.161
KTTC_GRN + RSR + NBR	0.243	79.5	+1.494	0.084	3	13	1	0.861	0.253
EVI + KTTC_GRN + RSR + NDII	0.269	79.5	+1.501	0.084	3	13	4	0.832	0.182
EVI + RSR + KTTC_WET	0.241	79.6	+1.668	0.077	3	13	2	0.851	0.227
RSR + NBR + KTTC_WET	0.239	79.8	+1.832	0.071	3	13	1	0.861	0.253
average model				**1.000**	**3**	**13**	**1**	**0.861**	**0.253**

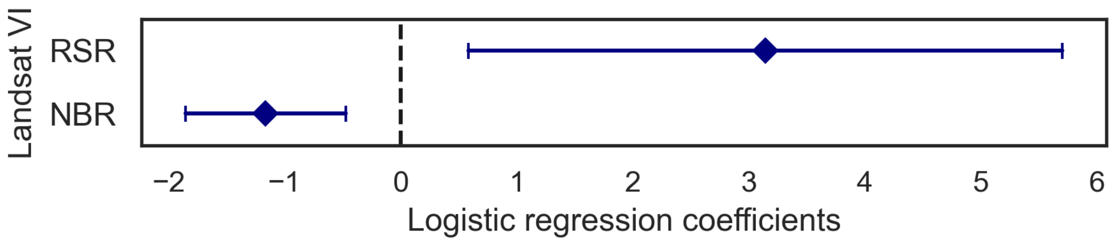

Figure S3: LMM logistic regression coefficients and 95% confidence intervals for the top (lowest *AICc*) unrestricted multi-VI model. See Table S4 for model accuracy metrics. Note that these coefficients differ little from the best result (that with the highest value of Cohen's κ; Figure S4) for the same VIs, except for the additional influence of GRVI in that result.

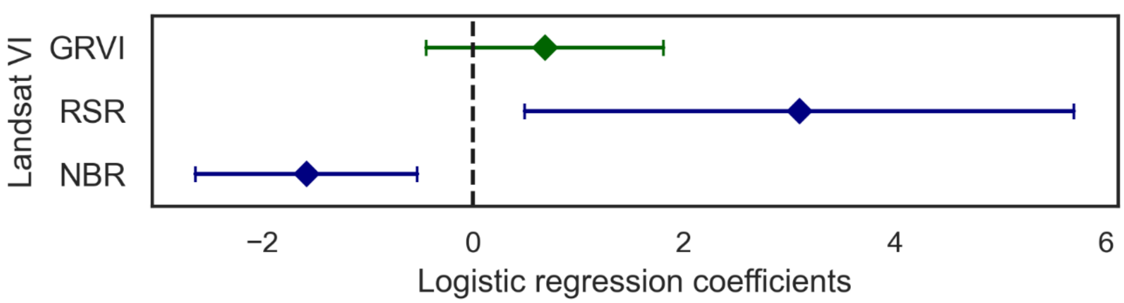

Figure S4: LMM logistic regression coefficients and 95% confidence intervals for the best unrestricted multi-VI model (that with the highest value of Cohen's κ). See Table S4 for model accuracy metrics and Table S5 for model prediction results.

Table S5: Results for mast-year prediction using the best unrestricted multi-VI logistic regression model based on GRVI, RSR, and NBR. Correctly predicted non-mast years (comprising 85 site-years) are not listed here. See Table S4 for model accuracy metrics and Figure S4 for logistic model coefficients.

Site	**Correctly predicted mast years (4)**	**Missed mast years (Type 2 error) (12)**	**False alarms (Type 1 error) (0)**
CHITTY	—	1998, 2010, 2014	—
KLOO	1993, 2010	1998, 2005, 2014	—
SILVER	2005	2010, 2014	—
SULPHUR	1993	1998, 2005, 2010, 2014	—