SCANFI version 2 update report (January 2026)

Natural Resources Canada | Ressources naturelles Canada

Luc Guindon[a], David L.P. Correia[a], Francis Manka[a] and Byron Smiley[b]

[a] *Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, 1055 rue du PEPS, Quebec City, QC, G1V 4C7, Canada*
[b] *Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, 506 Burnside Road, West, Victoria, BC V8Z 1M5, Canada*

This report provides a detailed description of all methodological and performance updates between version 1.1 of the Spatialized Canadian National Forest Inventory (SCANFI), originally published by Guindon et al. (2024), and version 2 of this data product.

SCANFI version 2 main updates

1. Transition to Landsat Collection 2

All Landsat inputs were updated from Collection 1 to Collection 2 summer surface reflectance (SR) and winter Top-of-Atmosphere (TOA) products. Landsat Collection 2 incorporates improved radiometric calibration, refined atmospheric correction, and updated geometric processing. For more information on Landsat Collection 2 refer to the USGS website.

2. SCANFI time series with LandTrendr-smoothed winter imagery

SCANFI version 1 relied on a 2020 3-year winter Landsat imagery median. While this methodology works well for 2020, it is unable to generate a reliable time series from 1985 onwards due to a lack of cloud-free winter imagery before 2000. SCANFI v2 now relies on Landsat TOA winter imagery processed with LandTrendr (Landsat-based Detection of Trends in Disturbance and Recovery; Kennedy et al. 2010), a temporal segmentation and smoothing algorithm that models pixel-level Landsat time series as a sequence of connected linear segments. This more complex data processing pipeline enables a 5-year interval time series from 1985 to 2025 with dual-season inputs. The resulting 5-year time series will be made publicly available.

3. Updated climate normals

The original McKenney et al. (2011) climate variables used in version 1 were replaced by the smoother 1990-2020 climate normals published by MacDonald et al. (2024). These new climate normals were modelled using more recent and denser weather station networks and modern interpolation approaches, which result in a spatially smoother product that eliminates the small tiling artifacts present in SCANFI v1. Additionally, by targeting the 1990-2020 time period, the new climate normals used in SCANFI V2 better represent contemporary climate conditions.

4. Increased sample selection density of NFI training data

SCANFI v1 sub-sampled photo plot data using a 250m systematic grid across all photo plots after removing all pixels that fell within 30m of photo-interpreted polygon boundaries. The SCANFI v2 sampling strategy still excludes pixels that are in proximity to polygon boundaries to avoid any co-registration issues, but relies on a denser sampling strategy, randomly sampling 4 pixels/ha per polygon, up to a maximum of 100 points per polygon (Fig. 1). As a result, the training set size increased from 124 287 to 588 570 data points, including a much larger proportion of all photo-interpreted polygons. The proportion of forested training points greatly increased, with mean crown closure going from 30.85% to 43.92% (Fig. 2). Only data from the National Forest Inventory (NFI) T1 photo plot measurement period, which covers data sampled between 2007 and 2017 were used. Data from the T2 remeasurement, which covers data from 2018 onwards, will be included in a later version once it becomes available.

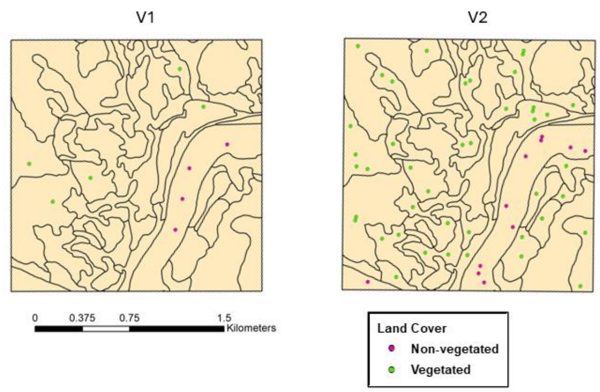

Figure 1. Comparison between SCANFI v1 (left) and v2 (right) sampling methodologies. Red points represent non-vegetated data points, and green points represent vegetated data points.

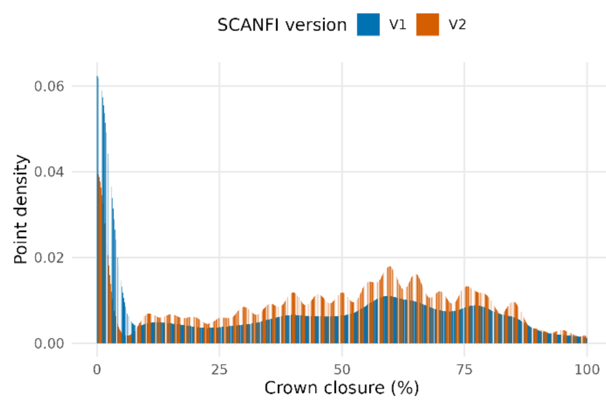

Figure 2. Density histogram of the crown closure distribution of SCANFI v1 (blue) and v2 (orange) training sets.

5. More accurate water prediction and delineation of urban areas and croplands

Water, urban, and cropland masking in SCANFI v2 is based on a novel, soon-to-be-published national land-use time series rather than a single static land-use product. This land-use layer is initialized using three complementary datasets to identify urban and cropland areas in 2025: the ESRI Land Cover World, the Agriculture and Agri-Food Canada (AAFC) land-use map (AAFC, 2020), and the Annual Crop Inventory (AAFC, 2024). From this 2025 baseline, annual forest area maps are used to determine land-cover states back to 1985, allowing urban and cropland classes to be dynamically converted back to forested land when forest cover exceeds 10%. This time-explicit approach prevents the misclassification of recently disturbed forests as water, an issue observed in SCANFI v1.0, and enables realistic reconstruction of forest extent prior to the availability of national AAFC land-use data. Together, these updates improve temporal consistency, reduce systematic land-cover errors, and yield more accurate forest, water, and other land cover delineation across the full SCANFI time series.

6. Forest age map

SCANFI v2 introduces a 1985–2025 time series of national forest age maps derived from a combination of three existing datasets: Canadian Landsat Disturbance 2 (CanLaD 2; Perbet et al., 2025), Pre-CanLaD (Correia et al., 2024), and SCANFI photo-interpreted age (Guindon et al., 2024). For recently disturbed stands, stand age was estimated as the time since the most recent stand-replacing disturbance using CanLaD 2 and Pre-CanLaD disturbance records. For undisturbed stands, stand age was derived from SCANFI-imputed National Forest Inventory (NFI) photo-interpreted age using a temporal harmonization procedure in which age estimates were generated at 5-year intervals, standardized to the target year, and aggregated using the median of all available estimates, thereby reducing noise relative to single-year predictions.

To generate the 5-year time series, the procedure begins with the 2025 age map and iteratively backtracks to 1985. At each step, disturbed areas are updated using the standardized median of the available pre-disturbance 5-year age estimates, while undisturbed areas are adjusted by removing the corresponding age difference to ensure temporal consistency. As a result, the most recent age maps are generally more accurate than earlier ones, since a larger proportion of disturbed areas are mapped using CanLaD 2 disturbance history records, which provide more precise disturbance timing than Pre-CanLaD and SCANFI age median alone. The age of water, urban areas, croplands, and other non-vegetated surfaces is set to NA.

7. Modified target and response variable sets

SCANFI v2 includes updates to both the modeled target variables and the predictor variable sets to improve thematic consistency and predictive performance.

7.1 Target variable updates

Water was removed as a modeled target variable and is now handled exclusively through the previously described time-explicit land-use layer, eliminating redundancy and reducing misclassification in disturbed areas. In addition, a new land position variable was introduced, distinguishing alpine, upland, and wetland environments, which considerably improved performance in alpine and wetland regions.

Table 1. Full list of spectral indices used to model SCANFI v2. Formula band nomenclature is as follows: R is red, G is green, B is blue, N is near infra-red, S1 is shortwave infra-red 1 and S2 is shortwave infra-red 2.

Season	Index	Formula	Reference
Summer	Automated Water Extraction Index	4.0 * (G - S1) - 0.25 * N + 2.75 * S2	https://doi.org/10.1016/j.rse.2013.08.029
	Chlorophyll Index Green	(N / G) - 1.0	https://doi.org/10.1078/0176-1617-00887
	Excess Green Index	2 * G - R - B	https://doi.org/10.13031/2013.27838
	Excess Red Index	1.3 * R - G	https://doi.org/10.1117/12.336896
	Green Ratio Vegetation Index	N/G	https://doi.org/10.2134/agronj2004.0314
	Modified Bare Soil Index	((S1 - S2 - N) / (S1 + S2 + N)) + 0.5	https://doi.org/10.3390/land10030231
	Modified Normalized Difference Vegetation Index	(N - S2) / (N + S2)	https://doi.org/10.1080/014311697216810
	Revised Multi-Spectral Water Index	-4.0 * ((B - G) / (B + G)) + 2.0 * ((G - N) / (G + N)) + 2.0 * ((G - S2) / (G + S2)) - ((G - S1) / (G + S1))	https://doi.org/10.3390/rs10101643
	Normalized Burn Ratio	(N - S2) / (N + S2)	https://doi.org/10.3133/ofr0211
	New Water Index	(B - (N + S1 + S2)) / (B + (N + S1 + S2))	https://doi.org/10.11873/j.issn.1004-0323.2009.2.167
Winter	Dry Bareness Index	((S1 − G) / (S1 + G)) − ((N − R) / (N + R))	https://doi.org/10.3390/land7030081
	Global Environment Monitoring Index	$((2 * (N^2 − R^2) + 1.5 * N + 0.5 * R) / (N + R + 0.5)) * (1 − 0.25 * ((2 * (N^2 − R^2) + 1.5 * N + 0.5 * R) / (N + R + 0.5))) − ((R − 0.125) / (1 − R))$	http://dx.doi.org/10.1007/bf00031911
	Normalized Difference Snow and Ice Index	(R − S1) / (R + S1)	https://doi.org/10.1080/01431160119766
	Non-Linear Vegetation Index	$(N^2 − R) / (N^2 + R)$	https://doi.org/10.1080/02757259409532252
	Normalized Green	G / (N + G + R)	https://doi.org/10.2134/agronj2004.0314
	Normalized Red	R / (N + G + R)	https://doi.org/10.2134/agronj2004.0314

7.2 Response variable updates

Topographic predictors were updated by replacing the ASTER digital elevation model (Abrams et al. 2020) used in SCANFI v1 with the recently published Canadian Medium Resolution Digital Elevation Model (Natural Resources Canada, 2024), providing improved elevation accuracy and spatial consistency. Landsat spectral predictors were comprehensively reviewed and optimized. First, a wide range of spectral indices were generated for both summer and winter imagery using the *rsi* R package (Mahoney, 2023). Highly collinear indices (Pearson correlation > 0.9) were removed to reduce redundancy. For each target variable, Recursive Feature Elimination (RFE) was then applied using the *caret* and *randomForestSRC* R packages (Kuhn, 2008; Ishwaran & Kogalur, 2025), considering only spectral bands and indices. The top four predictors per target variable were retained, and the original SCANFI v1 spectral feature set was fully replaced with this optimized selection. Selected spectral bands include summer Landsat bands 2-5 and 7, as well as winter Landsat bands 2, 3 and 5, according to the Landsat 7 nomenclature. Selected spectral indices are shown in Table 1.

8. Improved extrapolation of arctic ecozones

In SCANFI v1, land cover in arctic ecozones, which are located outside the NFI photo plot population (i.e. non-boreal and non-arctic ecozones), was predicted using a single national random forest model, while the same tile-level modeling framework was applied to predict all remaining forest attributes across these northern regions. Because these arctic areas fall well outside the environmental and structural conditions represented in the training data, this approach introduced tile-edge effects and spatial inconsistencies, particularly along transitions between tiles.

In SCANFI v2, this limitation was addressed by replacing the single extrapolation strategy with two region-specific models tailored to northern environments: (1) northern Quebec, 1504 data points; and (2) northern Northwest Territories and Nunavut, 2443 data points. Each model is trained using the northernmost available NFI photo plots from the corresponding geographic region, ensuring that training data more closely reflect local climatic, spectral, and ecological conditions. This regionally stratified approach substantially reduces tile-level artifacts, improves extrapolation stability, and produces a more spatially consistent and ecologically realistic product across Canada’s northern forests and arctic transition zones.

9. Updated time-since-disturbance layers

Integrated complete pre-1985 harvest and fire records (Correia et al. 2024), improving disturbance history initialization and enhancing stand-structure attribution. This dataset, along with its recently published update, identified over 30 million hectares that burnt or were harvest between 1950 and 1984. These disturbances are now taken into account and have a notable impact in the beginning of the SCANFI time series. All areas burnt and harvest between 1985 and 2025 are also considered thanks to the recently published CanLaD 2 (Perbet et al., 2025).

10. Modified tree species cover layers

In order to avoid interpretation errors and facilitate user analyses, SCANFI V2 tree species layers now represent the corresponding tree species crown closure, instead of the proportion of the overall crown closure that is represented by that species. Unlike SCANFI V1, users no longer need to multiply overall crown closure by tree species proportion (and divide by 100) to obtain tree species crown closure. Examples include:

- A pixel with 30% total crown closure composed of 50% black spruce and 50% jack pine will have 15% black spruce crown closure and 15% jack pine crown closure in SCANFI V2.
- A pixel with 40% total crown closure entirely dominated by balsam fir will have 40% balsam fir crown closure in SCANFI V2.
- A pixel with 100% total crown closure composed of 60% black spruce and 40% balsam fire will have 60% black spruce and 40% balsam fir crown closure in SCANFI V2.

Tree species aboveground biomass can still be obtained by estimating the proportion of the corresponding tree species relative to pixel-level crown closure and multiplying that value by pixel-level biomass.

Comparison between SCANFI version 1 and version 2 performance metrics

External validation performance

SCANFI v2 shows clear and consistent improvements over SCANFI v1 across all external validation datasets used to evaluate real-world performance in Guindon et al. (2024). Structural attributes validated with GEDI, Potapov, and airborne lidar show consistent gains in R^2 (approximately 2–3 points) and modest reductions in Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), indicating more accurate height and crown-closure estimates (Table 2). Species composition models evaluated with Quebec's Ministry of Forest, Fauna and Parks (MFFP) and the NFI ground plots (NFI 2021) exhibit the largest improvements, with R^2 increases ranging from roughly 4 to more than 15 points for several key species, including black spruce (*Picea mariana*), jack pine (*Pinus banksiana*), tamarack (*Larix laricina*), and broadleaves (Table 2). Biomass estimates show smaller but consistent

Table 2. External validation performance metrics of SCANFI version 1 and 2 across all external validation datasets used in the main manuscript. Best performing metrics are highlighted in light green.

Dataset	Points	Attribute	R^2		RMSE		MAE	
			V1	V2	V1	V2	V1	V2
GEDI	4790367	**Height**	50.14	52.95	6.26	6.12	4.06	3.95
Potapov	4620427	**Height**	84.07	85.68	3.51	3.58	2.08	2.15
Airborne Lidar	248642	**Height**	73.77	76.39	3.59	3.47	2.43	2.41
Airborne Lidar	248642	**Crown closure**	71.15	72.53	30.1	28.55	25.47	24.61
MFFP ground plots	2749	**Biomass**	35.56	40.87	44.65	41.91	34.32	32.32
		Balsam fir	38.75	48.59	22.62	20.84	14.91	13.72
		Black spruce	57.05	67.86	25.27	21.87	16.22	14.06
		Broadleaf	75.55	79.28	18.17	16.57	12.44	10.97
		Jack pine	30	45.39	16.75	14.78	7.66	6.93
		Tamarack	10.79	20.25	5.74	5.35	1.46	1.66
NFI ground plots	525	**Biomass**	41.26	42.89	67.25	66.22	47.34	45.43
		Balsam fir	28.86	24.77	12.81	13.18	5.98	6.00
		Black spruce	56.56	61.54	26.81	25.07	16.48	15.56
		Broadleaf	64.43	70.92	21.66	19.45	13.7	11.75
		Jack pine	27.08	31.04	14.58	14.58	6.54	6.86
		Tamarack	24.14	36.58	12.02	10.97	4.42	4.2
		Douglas fir	68.66	70.51	8.89	8.63	1.91	1.9
		Lodgepole pine	54.88	69.32	10.26	7.87	3.02	2.12

gains in explained variance and reduced error across both ground-plot datasets (Table 2). Overall, the results confirm that SCANFI v2 provides more accurate and robust predictions across a wide range of forest attributes and independent validation sources.

Internal cross-validation

Across all attributes, internal cross-validation metrics indicate slightly lower apparent performance for SCANFI v2 relative to SCANFI v1 (Table 3). For structural variables (biomass, height and crown closure), v2 shows modest reductions in R^2 and increases in RMSE, while relative error metrics (RMSE.r) are improved (Table 3). Species composition results are mixed, with v2 improving or matching performance for several species (e.g., black spruce, tamarack, white and red pine, other conifers) but showing lower R^2 for some less common species (e.g., lodgepole pine, Douglas fir, ponderosa pine). Bias remains low for both versions, with relative bias values close to zero across all attributes, indicating no systematic over- or under-prediction (Table 3). Overall, the table confirms that SCANFI v2 internal validation is more conservative, reflecting a more complex and representative training dataset rather than a degradation in external predictive performance.

Table 3. Leave-one-plot-out cross-validation performance metrics of SCANFI version 1 and 2. *RMSE.r* and *Bias.r* represent the corresponding performance metrics relative to the mean. Best performing metrics are highlighted in light green.

Attribute	R^2		RMSE		RMSE.r		Bias		Bias.r	
	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
Biomass	0.76	0.72	38.70	44.35	0.71	0.66	0.25	0.65	0.01	0.01
Height	0.78	0.75	4.28	4.63	0.56	0.48	0.05	0.07	0.01	0.01
Crown closure	0.82	0.76	13.94	15.18	0.45	0.40	0.09	0.37	0.003	0.01
Broadleaf	0.70	0.70	19.82	18.77	0.80	0.83	0.26	0.32	0.01	0.01
Jack pine	0.43	0.43	17.07	16.07	2.17	2.17	0.31	0.30	0.04	0.04
Lodgepole pine	0.58	0.50	14.95	14.31	1.86	2.31	0.19	0.22	0.02	0.04
Balsam fir	0.39	0.39	10.20	10.67	2.44	2.39	0.03	0.08	0.01	0.02
Douglas fir	0.60	0.55	9.10	8.65	3.07	3.70	0.04	0.05	0.01	0.02
White and red pine	0.28	0.37	3.92	4.13	7.97	6.82	0.01	0.01	0.01	0.02
Black spruce	0.53	0.57	25.64	25.21	0.95	0.83	0.43	0.61	0.02	0.02
Ponderosa pine	0.36	0.16	2.04	2.00	15.69	23.94	0.01	0.01	0.10	0.11
Other conifers	0.56	0.60	22.03	22.35	1.11	1.04	0.43	0.50	0.02	0.02
Tamarack	0.30	0.31	13.78	12.74	2.94	2.79	0.26	0.15	0.05	0.03

New SCANFI 5-year time series

SCANFI v2 includes a publicly available 5-year national time series (1985-2025) derived from LandTrendr-smoothed summer and winter Landsat imagery. The time series is built using cross-sensor harmonization to minimize artificial spectral shifts associated with changes in Landsat sensors, enabling consistent multi-decadal mapping of forest attributes. Despite this harmonization, some pixel-level temporal inconsistencies remain and have been intentionally preserved rather than smoothed out with additional post-processing algorithms to maintain SCANFI traceability for all predictions and ensure internal consistency across all predicted attributes.

At the pixel level, year-to-year variability can occur as a consequence of the k-nearest neighbour (kNN) framework used in SCANFI, which averages values from the three closest neighbours. Small changes in similarity between years may lead to localized temporal noise, such as minor tree species cover changes, or small gains and losses in predicted height between years. As a result, the SCANFI time series is not suitable for small-scale or plot-level temporal analyses, but is appropriate for large-area, regional, or national assessments, where this noise averages out spatially.

The SCANFI time series was also not specifically designed for biomass trend analysis. Predicting multiple forest attributes simultaneously generally results in lower accuracy than single-variable modeling approaches. Nonetheless, the observed national-scale aboveground biomass increase of approximately 4.8% (Figure 3) between 1985 and 2025 suggests that the time series remains within a plausible ecological range. Further targeted evaluation is required before using SCANFI for detailed biomass change attribution or carbon accounting applications.

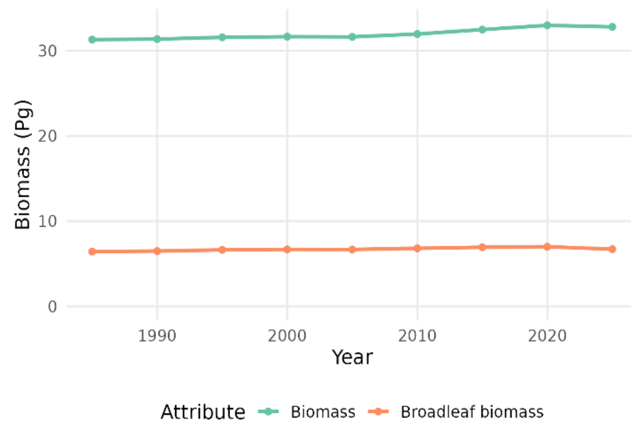

Figure 3. Sum of total live aboveground biomass (red) and total live aboveground broadleaf biomass (blue), in petagrams, across all Canada between 1985 and 2025.

Dataset citation

Guindon L., Correia D.L.P, Manka F. and Smiley B. 2026. SCANFI v2: Spatialized CAnadian National Forest Inventory data product v2. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Quebec, Canada. https://doi.org/10.23687/07653869-f303-46c2-a04e-9ab479b73cbf

References

Abrams, M., Crippen, R. and Fujisada, H., 2020. ASTER global digital elevation model (GDEM) and ASTER global water body dataset (ASTWBD). Remote Sensing, 12(7), p.1156.

Agriculture and Agri-Food Canada, 2020. 2020 AAFC land use map. Retrieved from open.canada.ca.

Agriculture and Agri-Food Canada, 2024. Annual Crop Inventory 2024 [Dataset]. Government of Canada. Retrieved from: https://agriculture.canada.ca/atlas/data_donnees/annualCropInventory/data_donnees/tif/2024/Correia, D.L., Guindon, L. and Parisien, M.A., 2024. Extending Canadian forest disturbance history maps prior to 1985. Ecosphere, 15(8), p.e4956.

Guindon, L., Manka, F., Correia, D.L., Villemaire, P., Smiley, B., Bernier, P., Gauthier, S., Beaudoin, A., Boucher, J. and Boulanger, Y., 2024. A new approach for Spatializing the

Canadian National Forest Inventory (SCANFI) using Landsat dense time series. Canadian Journal of Forest Research, 54(7), pp.793-815.

Kennedy, R.E., Yang, Z. and Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr-Temporal segmentation algorithms. Remote Sensing of Environment, 114(12), pp.2897-2910.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1-26. https://doi.org/10.18637/jss.v028.i05

Ishwaran, H. and Kogalur, U.B., 2025. Fast Unified Random Forests for Survival, Regression, and Classification RF-SRC), R package version 3.4.1.

MacDonald, H., McKenney, D.W., Pedlar, J., Lawrence, K., de Boer, K. and Hutchinson, M.F., 2024. Spatial datasets of 30-year (1991–2020) average monthly total precipitation and minimum/maximum temperature for Canada and the United States. Data in Brief, 55, p.110561.

Mahoney, M., 2023. rsi: Efficiently Retrieve and Process Satellite Imagery. doi:10.5281/zenodo.10926857 <https://permian-global-research.github.io/rsi/>.

McKenney, D.W., Hutchinson, M.F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R.F., Price, D. and Owen, T., 2011. Customized spatial climate models for North America. Bulletin of the American Meteorological Society, 92(12), pp.1611-1622.

Natural Resources Canada, 2025. Medium Resolution Digital Elevation Model (MRDEM) – CanElevation Series – Product specification. Edition 1.2 (2025). Retrieved from open.canada.ca.

NFI, 2021. National Standards for Ground Plots Compilation Procedures, version 2.4. Available from https://nfi.nfis.org/en/ground_plot [accessed 29 Decembre 2025]

Perbet, P., Guindon, L., Correia, D.L., Gahrouei, O.R., Côté, J.F. and Béland, M., 2025. Historical insect disturbance maps from 1985 onwards for Canadian forests derived using earth observation data. Scientific Data.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E. and Armston, J., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment, 253, p.112165.