

# **REGENERATION IMPUTATION MODELS FOR INTERIOR CEDAR HEMLOCK STANDS**

**Badre Tameme Hassani, M.Sc.,**

**Peter Marshall PhD., Valerie LeMay, PhD., Temesgen  
Hailemariam, PhD., and Abdel-Azim Zumrawi, PhD.**

**Originally Presented at Western Mensurationists Meeting,**

**June 23-25, 2002, Leavenworth, WA**

**(Presented Again at the Regeneration Modelling Workshop, Dec. 17, 2002, UBC)**

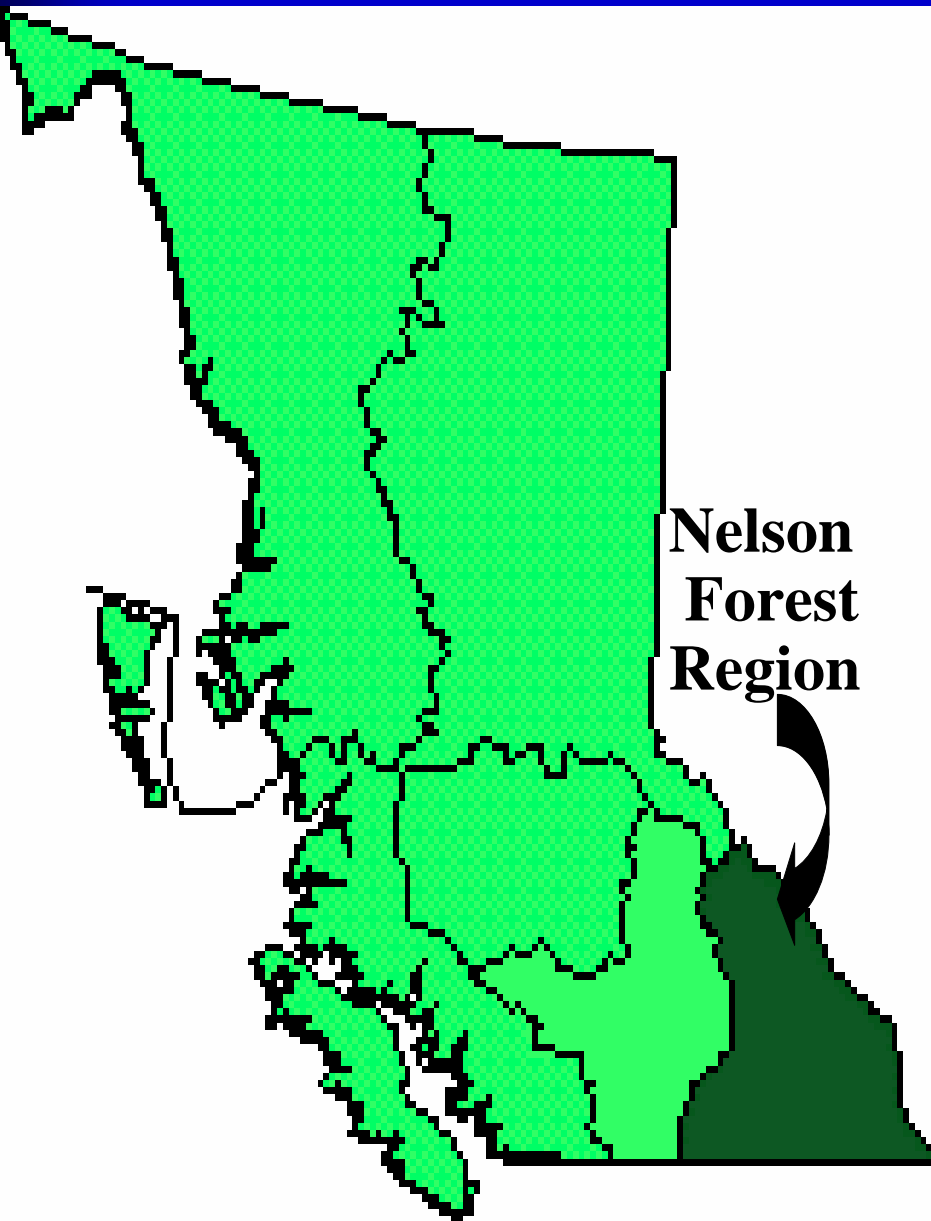
# Background

- Understanding stand dynamics is necessary to achieve management objectives
- Regeneration is the earliest stage of stand development
- Prognosis<sup>BC</sup> has been calibrated by the MoF for use in southeastern portion of BC
- Regression approaches did not lead to good predictions of regeneration
- Currently, the regeneration portion of Prognosis<sup>BC</sup> has been disabled

# Objectives

- Explore the use of imputation techniques to predict regeneration in the complex mixed-species stands of Interior Cedar Hemlock zone
- Predict regeneration using some of the imputation methods

# Location of Study Area in BC



ICHmw2

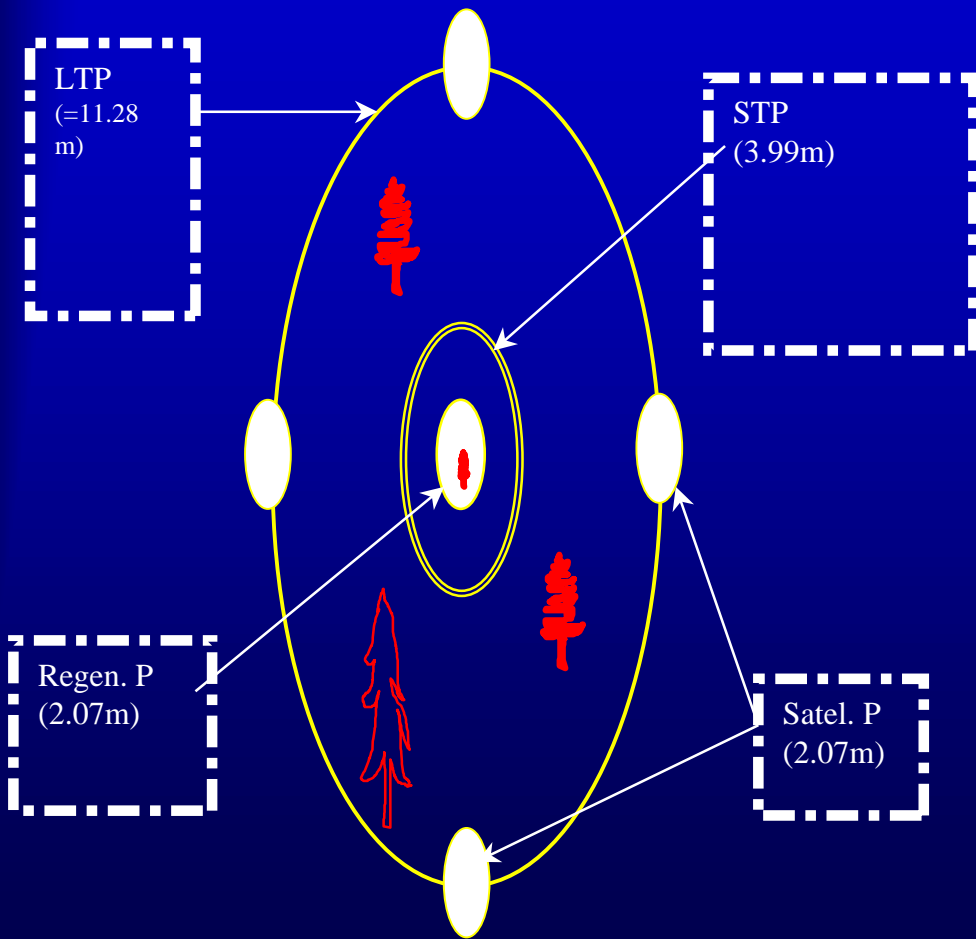
- Continental climate
- Lower to middle elev.
- Most productive in the interior of BC
- Supports 15 trees species

# Complex Stands in ICH zone

- Mixed species
- Uneven aged
- Multi-cohort
- Cedar and hemlock are climax species



# Plot Layout



- Nested plot
- Systematic location in Selected stands
- Stands selected to cover the range of:
  - overstory density
  - age since disturbance
  - site preparation
  - Slope percent
  - Aspect
  - elevation
- 333 Plots from 138 Polygons

# Species Groups

---

## Species Group

## Species

---

Shade tolerant

Cedar, hemlock, grand fir, subalpine fir, spruce

Shade semi-tolerant

Douglas-fir, white pine

Shade intolerant

Larch, lodgepole pine

Hardwood

Aspen, cottonwood, birch, Douglas Maple, willow, yew<sup>1</sup>

---

<sup>1</sup> western yew (Tc) is a coniferous species, rare and not commercial

# Tabular Imputation Approach

## Stand Conditions:

- **Basal Area class** (Dense:  $> 5\text{m}^2/\text{ha}$ ; Open:  $\leq 5\text{m}^2/\text{ha}$ )
- **Site Series class** (Dry: 02-03, Slightly Dry: 04, Mesic: 01, Slightly Wet: 05, and Wet: 06-07-08)
- **Time-since-disturbance class (years)**: (1: 1-5, 2: 6-10, 3: 11-15, 4: 16-20, and 5: 21-25)

# Tabular Imputation Approach

For each stand condition combination (using all data):

- Average number of seedlings per ha by:
  - height class (1: 15-49.9 cm, 2: 50-99.9 cm, 3: 100-129.9 cm, and 4: >130 cm)
  - **and** for the 4 species groups (16 regeneration variables)
- Sample Statistics for each cell (species and height):
  - Standard error of the mean
  - standard deviations
  - coefficients of variation

# Testing of the Tabular Imputation Model

## Validation:

- Data randomly split into 5 subsets (20%)
- each subset was set aside once for model evaluation
- Calculate the root mean squared error (RMSE) and RMSE/mean observed values by plot
- Also looked at model accuracy within the 16 cells (4 heights by 4 species groups)

# Model accuracy over cells

- Match: Presence of regeneration in both the observed and expected cell (4 species \* 4 height classes = 16 cells)
- Classified predicted plots into:
  - good (>90% matched),
  - moderate (50%-90% matched), and
  - poor (<50% matched) classes
- For each class, grouped plots also by RMSE:
  - low (<1000),
  - moderate (1000-2000), and
  - high (>2000) RMSE

# Most Similar Neighbour (MSN)

- Find a similar polygon from a set of **reference** plots (have detailed information) and use the data from the substitute for the **target** plot
- Retains the variability of the variables over the stand (forest, landscape), as represented by the reference polygons

# Distance metric to select neighbours

Most Similar Neighbour

$$d_{ij}^2 = (X_i - X_j)' \Gamma \Lambda^2 \Gamma' (X_i - X_j) \quad [1]$$

$X_i$  : vector of standardized values,  $i^{\text{th}}$  target plot

$X_j$  : vector of standardized values,  $j^{\text{th}}$  reference plot

$\Gamma$  : matrix of standardized canonical coefficients for the X variables

$\Lambda^2$  : diagonal matrix of squared canonical correlations

# MSN Approach

**Three MSN analyses were conducted:**

- **MSN Type 1:** Number of seedlings per ha by 4 height classes for the 4 species groups
- **MSN Type 2:** Number of seedlings per ha by 2 height class (0.15 to 1.30 m and > 1.3m) for the 4 species groups
- **MSN Type 3:** Number of seedlings per ha for the 4 species groups (without height class)

# Variables for MSN

- Auxiliary variables (X set):
  - 8 continuous variables: Years since disturbance, site series, elevation, elevation, slope percent, basal area /ha, and CCF
  - and 2 class variables (Slope position (5), and site preparation (5)).
- Regeneration per ha variables (Y set):
  - 16, 8, or 4 variables depending on the MSN type

# Data and Validation

- Same data as used for the tabular imputation approach
- Data randomly split into 5 subsets (20%)
- In each of the 5 runs, one subset represented target plots (assumed to lack regeneration information) and the remaining 80% represented reference plots (have complete information)

# Testing of the MSN

- The 3 types of MSN were compared using
  - bias (mean deviation),
  - mean absolute deviation, and
  - RMSE
- For the best MSN type only, observed and the predicted regeneration of target plots were compared using combinations of:
  - the number of matched categories
  - and RMSE

# Results

## Tabular Imputation Method

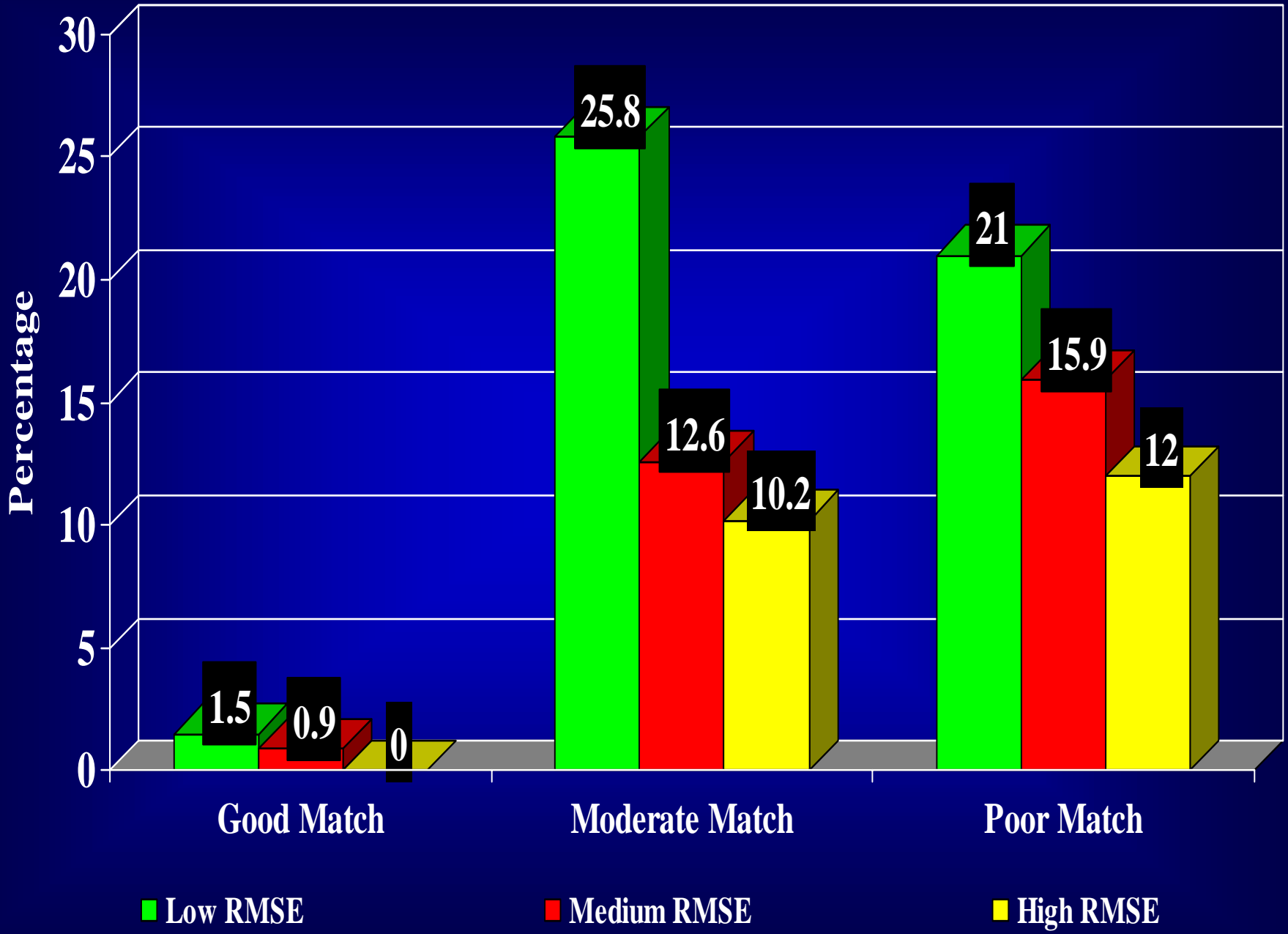
- 50 tables were produced

# Dense, Dry, first 5 years since disturbance (n=18)

Species	Height (cm)				Total
	15-49.9	50-99.9	100-129.2	>130	
<b>Tolerant</b>	3921	1032	454	495	5903
<b>Semi-toler.</b>	2889	949	372	578	4788
<b>Intolerant</b>	1197	41	41	0	1280
<b>Hardwood</b>	454	248	248	743	1692
<b>Total</b>	8462	2270	1115	1816	13663

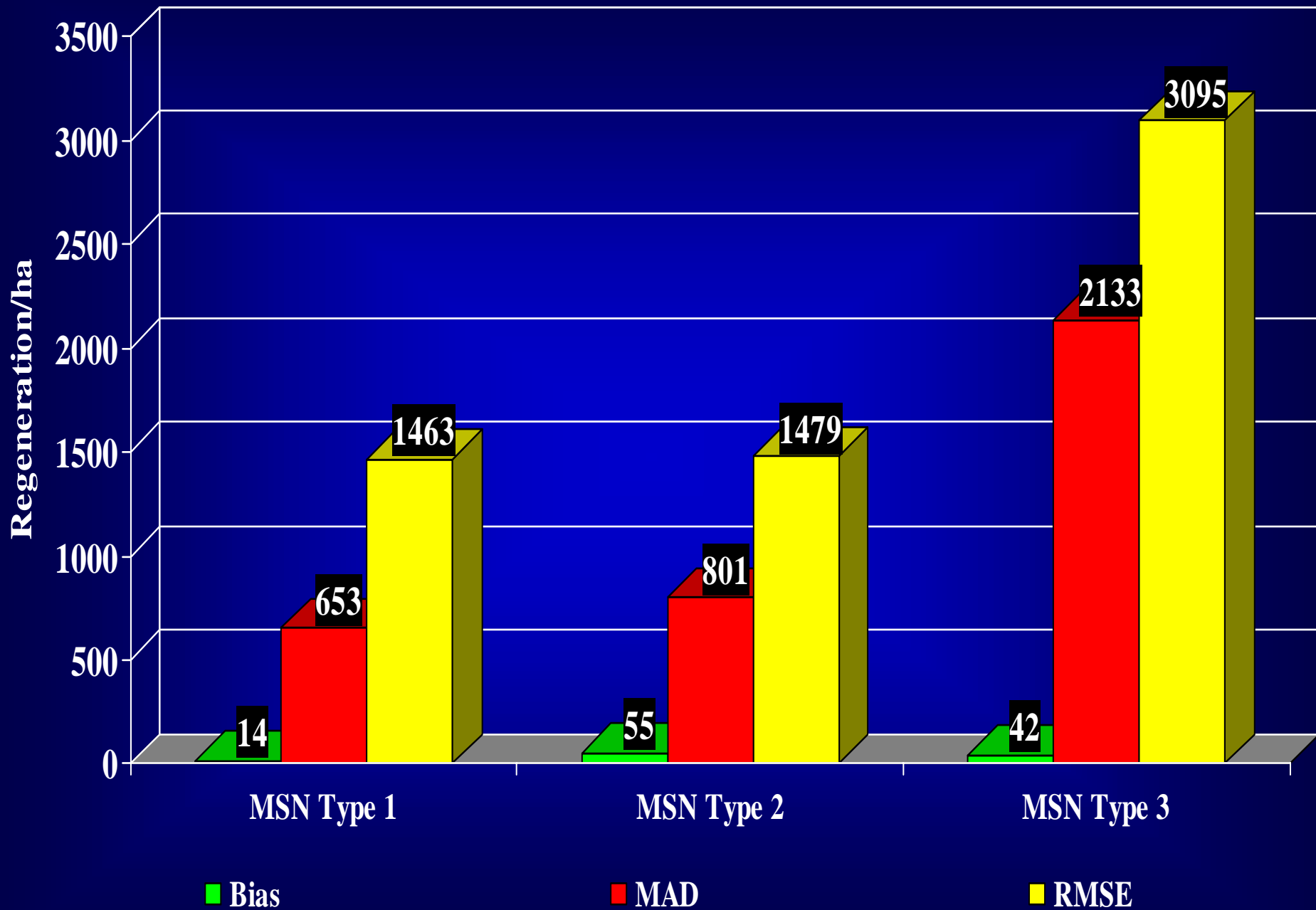
# Validation of Tabular Imputation Models

- Predictions based on less than 10 plots resulted in very high standard errors of the mean (SEE) (reaching 500 % of the mean in some cases)
- Predictions based on between 10 and 20 plots showed a slight decrease in SEE
- No obvious trend over age since disturbance was evident across any stand condition

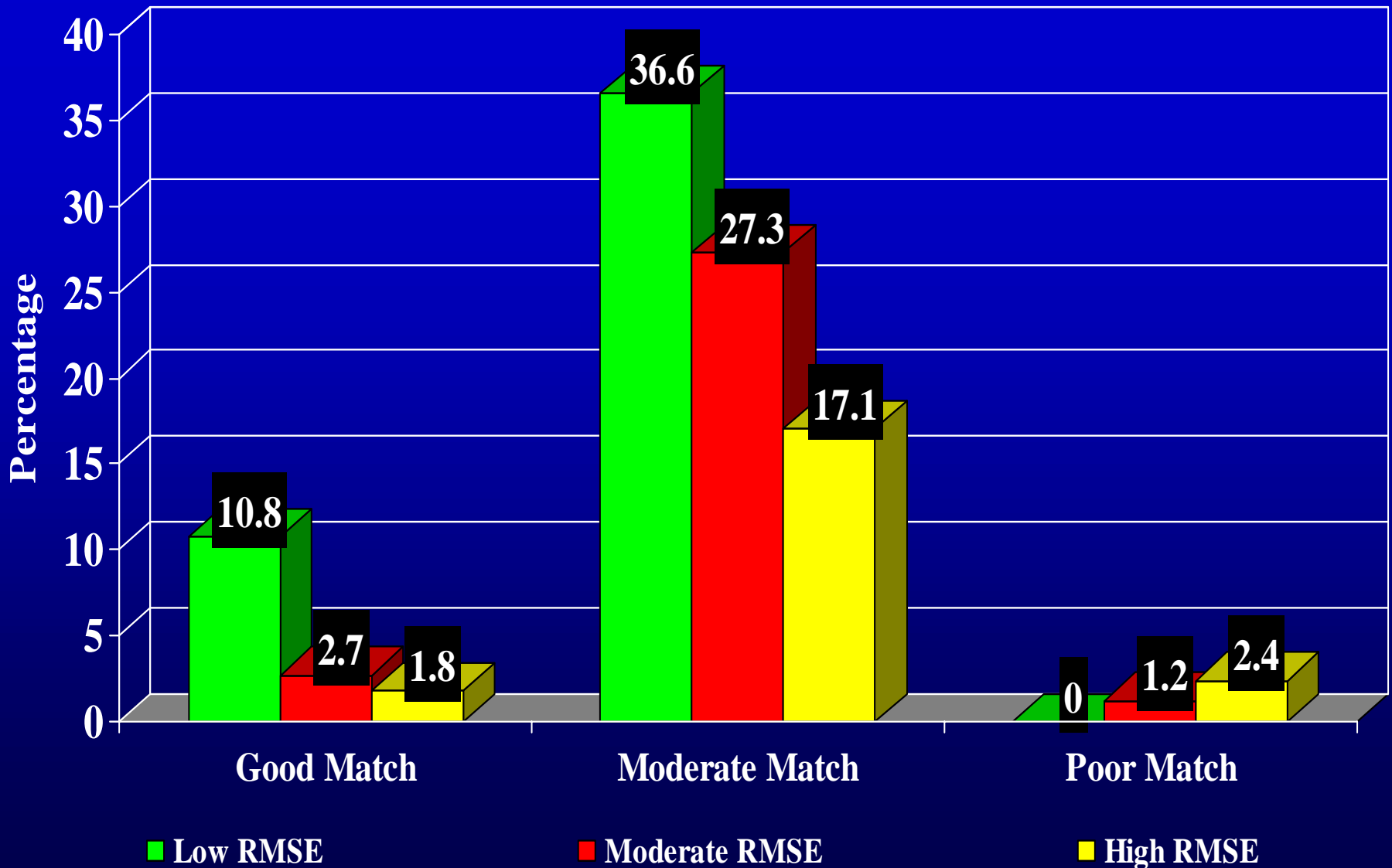


# MSN Approach

- Low correlations between the regeneration (Ys) and the auxiliary variables (Xs)
- Stand density indicators (Basal area, Trees per ha, and CCF) had the highest correlation coefficients



# Performance of MSN Type 1



# Comparison of Approaches

Run	# of target plots	Type 1 (4 height classes*4 species groups) Model					
		MSN			Tabular		
		<i>Bias</i>	<i>MAD</i>	<i>RMSE</i>	<i>Bias</i>	<i>MAD</i>	<i>RMSE</i>
<b>1</b>	68	36	638	1,432	-110	698	1,472
<b>2</b>	65	-35	666	1,547	-188	705	1,407
<b>3</b>	65	154	576	1,214	-144	773	1,578
<b>4</b>	64	-143	792	1,802	-99	660	1,298
<b>5</b>	71	58	594	1,323	-200	813	1,701
<b>Mean</b>	<b>66.6</b>	<b>14</b>	<b>653</b>	<b>1,463</b>	<b>-148</b>	<b>730</b>	<b>1,491</b>

# Conclusions

- Performances of the imputation techniques depend implicitly on the data used in the analysis
- Both approaches were successful in predicting regeneration by making use of available data
- Tabular approach had a simple structure, provided realistic and detailed postharvest regeneration sites
- The MSN approach was robust, more flexible, and was a better predictor than the tabular approach

# Conclusions

- Separating advance and subsequent regeneration would possibly improve imputation predictions even more successful
- Could explore the applicability of other imputation techniques, such as K-NN (k-nearest neighbour), to predict regeneration might improve the accuracy of regeneration estimates

# Acknowledgments

- This research was funded by the Resource Inventory Branch, Research Branch, and Forest Practices Branch of the BC Ministry of Forests via FRBC funding