Avian Database v4 - Manual

Someone, someone

October 2018

Contents

---

1 INTRODUCTION

The boreal region of Canada hosts one of the most diverse bird communities in North America, comprised of more than 300 species. During the summer, more than 300 bird species and up to three billion individual birds breed in the region, leading to its characterization as a North American bird nursery. Historically, these forests have been subject to little widespread development pressure, and access has been extremely limited, resulting in a corresponding gap in research and monitoring. However, pressures are mounting, and with the rapid development of the boreal forest that is presently occurring, there is an urgent need to understand the impact of changing habitats on boreal bird populations in order to inform management actions. There has been a recent surge of boreal bird studies across Canada, thus it is timely to undertake a synthesis of existing research, and as a first step towards scenario analysis and decision-support, formulate habitat-based predictive models of species abundance and distribution for Canadian boreal forests. We envision a series of spatially-explicit, bird-habitat models, broadly accessible to all organizations interested in boreal conservation planning, and updated regularly with monitoring data and new research results. The objective of the BAM

Avian DB is to assemble and harmonize existing data sets on all boreal forest birds and their habitats through cooperative efforts with boreal bird researchers and associated agencies.

2 BAM AVIAN DATABASE

The Boreal Avian Modelling (BAM) Project has been assembling a comprehensive database of systematic, spatially referenced, observational data on boreal birds that have been collected in the boreal and hemiboreal regions of Canada and the United States. Our primary focus is point-count data. This standardized survey technique, used widely in avian research, is well suited for evaluating habitat requirements of forest songbirds.

To date, we have compiled much of the point-count data collected in the Canadian boreal forest since 1993. The avian data set contains information from more than 135 individual research projects comprising more than 100 000 off-road point-count locations, and over 1 million bird records. Approximately 100 bird species are represented. Recent data compilation efforts have resulted in the acquisition of the majority of geo-referenced point-count survey data from across boreal and Arctic Alaska. Efforts continue to acquire data from the hemiboreal region across North America including the Upper Midwest and New England areas of the US.

2.1 Database Structure

The Avian Database contains the observational survey data contributed from all projects, standardized into a common format and ready to use in analyses. The observational data and covariates are maintained in a relational database to facilitate data handling by users. It consists of four hierarchical normalized tables - Projects metadata, Station locations, Sampling events and Point Count data- that are all connected using unique key identifier (Figure 1).

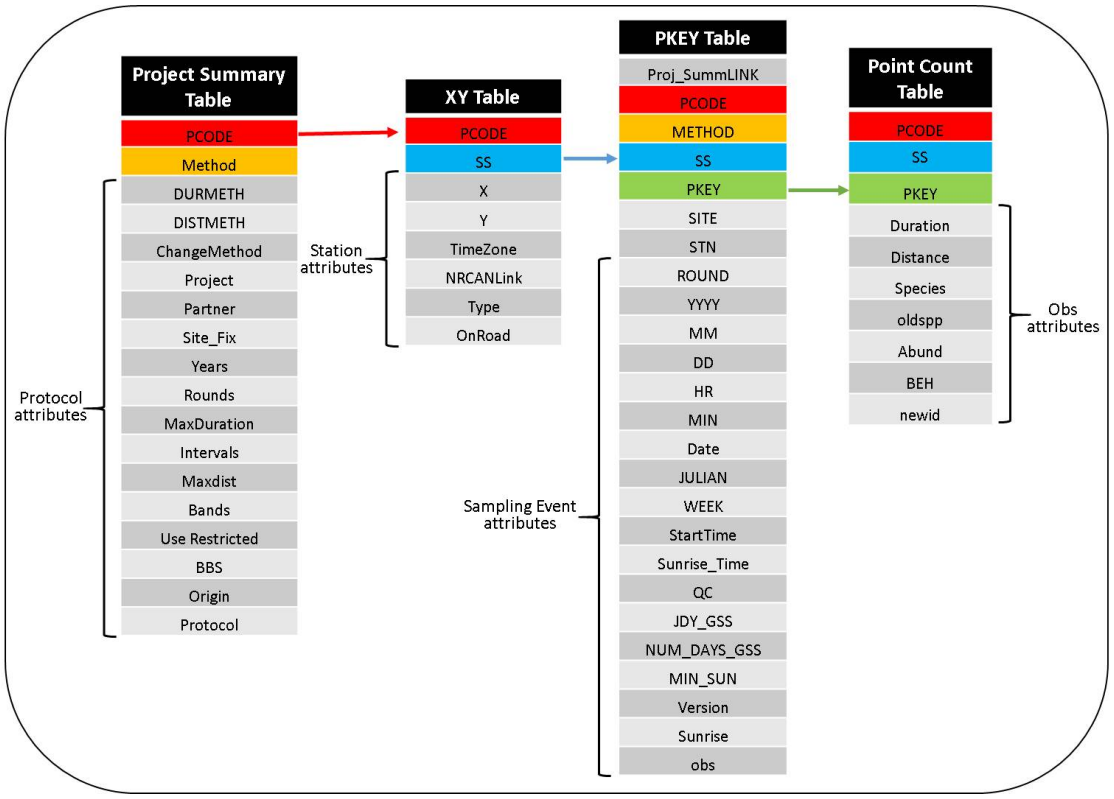

Figure 1: Relational Boreal Avian database schema

Description of the four tables is listed below:

1. Project Summary Table – Preserves essential metadata for each project, including sampling protocol employed, and any restrictions on data use;
2. XY Table – value that pertain geographic coordinates for each survey location;
3. PKEY Table – values that pertain to data and time of visit to a particular survey location, as well as observer identity if available and included;
4. Point Count Table – values that pertain the actual survey information, such as the number of individuals of each species detected, any information about distance to the detected species, the time of detection relative to survey initiation, and available information on the bird's behavior.

The main body of this report defines each of the four tables and tabulates the attributes and their characteristics. A summary of the data structure and data dictionary is also presented for each table. Maintaining this information within the database ensures consistent classifications across users, and prevents repetition of background searches on life history information each time a new analysis is undertaken.

2.2 Project Summary Table

The BAM database contains contributed data from 135 distinct projects. Projects are sets of data contributed by a single owner or entity. In principle, all data within a single contributed project were collected under a common survey protocol. The two most important elements of

a survey protocol are the sampling duration (DURMET) (e.g. 5, 10) and the sampling distance or radius (DISTMET) (e.g. 50m, 100m, unlimited distance) (Table 1). Other components of a sampling protocol include the number of years over which the study was conducted, and the number of visits per year. These characteristics are not always uniform for all surveyed locations within a project. The field Method allow to distinguish within a project difference in sampling duration and distance.

Table 1: Project Summary Table Structure

Field Name	Fields Description	Data Type	Code Description
PCODE	Project code. Allow to link data sharing agreement to stations	Character	Section 2.2.1, Table 2
Method	Unique project identification code. It is the concatenation of PCODE with a unique identifier. It allows to link Project metadata to Point Count data according to the methodology used during the survey	Character	
DURMETH	Duration methodology used during the survey	Character	Section 2.2.2, Table 3
DISTMETH	Distance methodology over which data were recorded during the survey	Character	Section 2.2.3, Table 4
ChangeMethod	Indicates if the methodology change at any point of the project (over years or within year). Most of the time, changes will be reflected by a change in Method field	Boolean	Table 5
Project	Project name from which the point count data came	Character	
Partner	Name of the contributor	Character	
Site_Fix	Allow to reassign sites for project where they were too dispersed (specifically created for Diana's analysis)	Boolean	Table 5
Years	Number of year sampled. It does not imply that each station was visited over repeated years	Character	
Rounds	Number of rounds sampled in a year	Character	
MaxDuration	Maximum time of individual count. It is directly derived from DURMETH	Character	
intervals	Number of time intervals used in count. For example, if the duration sampled is 0-3min, 3-5min, number of intervals = 2	Character	
MaxDist	Maximum distance sampled of individual count in the project methodology. It is directly derived from DISTMETH	Character	
Bands	Number of distance bands in sampling method. For example, if bird count were recorded within a radius classes 0-25m, 25-50m, 5-100m, bands = 3	Character	
Use Restricted	Identifies whether the use of the data is unrestricted, restricted or limited	Boolean	Table 5
BBS	Projects where protocols are BBS specific but not bbs data	Boolean	Table 5
Origin	Where the bulk of a dataset was being pulled from - also the Atlas data is a bit separate, the ALMS data is a huge amount of data from Alaska	Character	
Protocol	Identifies if projects protocols was derived from another project (i.e.; ATLAS)	Character	

2.2.1 PCODE

Each Project is assigned a unique code (PCODE) which links to data sharing agreements, and also serves as a key to relate project information to stations (Table 2). A single project can however use different methodology. To link project information to observations found in PointCount Table, use Method field.

Table 2: List of PCODE used in BAM Avian DB v4

PCODE	Stations	Year	TotalVisits
ABCAWAWEST	1177	1	1177
AD02	40	1	40
AD04	102	1	203
AD1	151	3	452
AK_YC	1415	2	1413
AKALMS_DIST	783	10	2249
AKCAFS	103	1	103
AKDEN	807	7	1325
AKOR	115	5	521
AKOR_CNT	2034	17	11530
ALMS	1130	8	2074
BCCA	10089	3	10089
BM	48	1	48
BNP	174	12	381
BR	88	3	556
CF	1040	4	2589
CHSS	40	1	77
CL	469	22	30525
CLSS	66	1	127
COCL	145	1	145
COMW	179	1	179
CVME	105	1	105
CW	577	2	577
DA	88	7	1038
DARV	145	3	648
DP	142	1	142
DRAP	208	1	416
DV	162	1	162
ECJOSM	4499	3	4499
ECJOSM_JRB	2490	1	2490
EHPP	51	2	52
EM	106	4	1554
EMB-ASP	670	2	670
EMB-BS	223	1	268
EMB-NOISE	168	1	670
EMCLA	555	2	2906
EMCLA2014	759	1	3280
FBMP	55	6	381
FFBR	630	2	1880
FG	233	4	417
FLPC	275	8	3531
FSLE	54	2	216
FTL	73	1	73
FtLiardRd	57	1	57
GLDRcnrlbl	21	1	21
GLDRCNRLEXP	64	1	64
GLDRCNRLHE	31	1	31
GLDRCNRLHZ	108	1	108
GLDRCNRLPE	77	1	77
GLDRMEGO	56	1	56
GLDROPTILL	160	1	160
GLDRPCCL01	69	1	69
GLDRPCCL03	67	1	67

Table 2: List of PCODE used in BAM Avian DB v4 *(continued)*

PCODE	Stations	Year	TotalVisits
GLDRPCMC	88	1	88
GLDRPCMCII	45	1	45
GLDRRAKL	112	1	112
GLDRSHL13E	89	1	89
GLDRSHMRM	113	1	113
GLDRSVEXP	90	1	90
GLDRSVSB	44	1	44
GLDRSVTP	23	1	23
GLDRSVUPG	27	1	27
GPMN	147	1	147
HEDW	385	1	385
HOBBBS	1925	2	3295
HR	572	2	1111
IMBE	113	4	460
ITHNT	57	1	57
JL	104	2	231
JNP	119	4	385
KENO	110	1	110
KH	58	1	58
KNP	83	2	129
KTPC	257	1	257
KX	73	1	73
Lebl	95	2	358
LMWELL	64	1	128
LP	1663	4	3320
LR	126	1	286
MANBBA	8345	2	8345
MAVEXP	64	1	127
MB06	404	1	404
MBBA	11810	5	11810
MC	110	3	110
MGLE	82	2	248
ML	102	1	200
MM94	347	2	730
MM95	243	1	243
MNBBA	7021	5	7027
MNNFB	1842	24	32036
MR	120	2	289
NLCTRLAB	191	1	191
NLLABHIGH	116	2	369
NLLR	32	1	32
NLMR	80	1	150
NWSS	59	1	107
ON	52904	5	52904
PA	382	2	765
PERI	109	1	109
PF	313	3	509
PN	446	1	446
PR	137	1	395
QCATLAS	12193	3	12193
RLMBP	472	2	569
ROMA	280	1	280
RP	2695	4	2705
RUEA	96	1	96
RUST	97	1	97
SC	187	1	187
SH	394	1	1177
SKAMP	108	2	108
SKBS	370	1	370
SKPC	229	2	229
SNPC	397	2	397

Table 2: List of PCODE used in BAM Avian DB v4 *(continued)*

PCODE	Stations	Year	TotalVisits
SRDR	306	1	306
THIN	70	6	2341
TTPC	227	2	227
WAP	539	3	3104
WBHS	80	1	80
WF	146	6	784
WH	616	2	1142
WLNP	118	4	407
WR	37	6	444
YK_CM	193	1	193
YKAGP	185	1	185
YKBL	10	1	10
YKDLUP	106	1	106
YKFL	156	1	155
YKGAP	62	1	123
YKKEMP	57	4	147
YKLIARD	125	1	245
YKRIPARIAN	203	1	322
YKSW	144	1	575
YKTeslin	269	6	1032
YNP	108	2	186

2.2.2 Durmeth

Durmeth is the duration protocol used during the survey. Observations were recorded using time interval classes. For example, in a 5 minutes point count survey, time intervals classes could be splitted into 0-3min and 3-5 minutes. Unit is in minutes (Table 3).

Table 3: List of Duration Methodology Class

Duration Code	Duration Range
A	0-10
B	0-5
C	0-5-10
D	0-3
E	0-20
F	0-5-10-10+
G	0-3-5-10-10+
H	0-3-10
I	0-3-5
J	Unknown
K	0-5-10-15-20
Q	0-10-15
R	0-3-5-8-10
S	0-5-8
T	0-3-7-10
X	0-3.33-6.66-10
Y	0-1-2-3-4-5-6-7-8-9-10
Z	0-2-3-4-5-6-7-8-9-10
AA	0-2
Ya	0-1-2-3 (4-5-6-7-8-9-10)
BB	0-1-2-3

2.2.3 Distmeth

Dismeth is the distance protocols over which data were recorded during a survey. Some protocols stratified observations using interval class (also called detection radius), such as less than 50m, 50-100m and greater than 100m. Distance unit is in meters (Table 4).

Table 4: List of Distance Methodology Class

DISTMETH	Distance Range
A	0-75-75+
B	0-50-100-100+
C	0-50-UNL
D	UNLIMITED
F	0-100
G	0-80
H	0-50-75-100-UNL
I	0-25-50-75-75+
J	0-60-60+
K	0-50
L	0-50-100-150
M	0-25-50-75-100-125-150-150+
N	0-100-100+
O	Unknown
P	0-30-50-75-
R	0-50-75-100
S	0-25-50-75-100-100+
T	0-50-100
U	0-10-20-30-40-50-60-70-80-90-100-125-150->150
V	0-10-25-50-100-100+
X	Recorder
W	0-10-20-30-40-50-60-70-80-90-100->100-125-150->150

2.2.4 ChangeMethod, Site_Fix, Use restricted and BBS

ChangeMethod, Site_Fix, Use restricted and BBS fields used Boolean data type where 1 means Yes and 0 means No (Table 5). For example, a project that has ChangeMethod = 0 indicates that the method didn't change at any point in the project (over years or within year).

Table 5: Boolean Code Description

Code	Description
0	No
1	Yes

2.3 XY Table

Stations are the geographically referenced locations where individual point counts were conducted. Each station is linked to a Project via PCODE,and the location is recorded in geographic coordinates. A unique key (SS) is assigned to each station. In constructing the key, we used as far as possible the project-specific naming conventions; this allows us preserve some features of the sampling designs for individual projects, and to identify groups of stations that should potentially be grouped in random effects models. We allow one level of grouping within projects, and stations are numbered within groups.

Table 6: XY Table Structure

Field Name	Fields Description	Data Type	Code Description
PCODE	Project code. Allow to link data sharing agreement to stations	Character	Section 2.2.1, Table 2
SS	Sampling site identifier. Concatenation of PCODE:SITE:STN	Character	
X	X coordinate using the Canadian Spatial Reference System NAD 83	Character	
Y	Y coordinate using Canadian Spatial Reference System NAD 83	Character	
TimeZone	Time zone where project took place	Character	Section 2.3.1, Table 7
NRCANlink	Link data to NRCAN weather data. No longer used	Character	
Type	Type of protocol. Duplicate from Project Summary Table protocol	Character	
OnRoad	Sample was done on a road	Boolean	Section 2.3.2, Table 8

There is also a unique numeric identifier (ID_link) assigned to each station, independent of the constructed key. This table may also include one or more records spatially locating the station within ecoregions, BCRs or other geographic zonations that can be coded as factors. However, such attributes are not guaranteed to be maintained across database versions, and should properly be located in a separate table linked by (SS).

2.3.1 Timezone

There are duplicates and typo. The DB should be checked and fixed prior to update csv

Table 7: TimeZone Code Description

TimeZone	Description
MDT	Mountain Daylight Time
EDT	Eastern Daylight Time
CDT	Central Daylight Time
AKST	Alaska Standard Time
PDT	Pacific Daylight Time
NDT	Newfoundland Daylight Time
MDT	Mountain Daylight Time
ADT	Atlantic Daylight Time
edt	Eastern Daylight Time

2.3.2 OnRoad

For now, I have no clue what those codes means. Should be confirmed by Trish

2.4 PKEY Table

PKEY table represents the individual point counts information such as date, time, observer, etc. Rounds record the year, calendar date and clock time when data were collected at a station. Clock time is time-zone corrected given the location and date of sampling. For stations with multiple visits within years, a Round number is also recorded. Dates are also

Table 8: OnRoad Code Description

OnRoad Code	Description
NULL	NA
NO	NA
2	NA
N	NA
Y	NA
UNK	NA
3	NA
Yes	NA
1	NA

recorded as Julian dates and relative to the start of the local growing season, as determined by custom interpolated climate data. Times are also recorded relative to local nautical(?) sunrise at the geographic coordinates and date of sampling. Each Visit is assigned a unique PKEY composed of SS, a two-character code for sampling year (e.g. 99 = 1999, 07 = 2007) and the Round.

Table 9: PKEY Table Structure

Field Name	Field Description	Data Type	Code Description
Proj_Summlink		character	
PCODE	Project code. Allow to link data sharing agreement to stations	character	Section 2.2.1, Table 2
METHOD	Unique project identification code. It is the concatenation of PCODE with a unique identifier. It allows to link Project metadata to Point Count data according to the methodology used during the survey	character	
SS	Sampling site identifier. Concatenation of PCODE:SITE:STN	character	
PKEY	Individual point count identifier. Concatenation of SS:YY:ROUND	character	
SITE	?	character	
STN	?	character	
ROUND	Number of visit within a year	character	
YYYYY	Year	integer	
MM	Month	integer	
DD	Day	integer	
HR	Hour	integer	
MIN	minutes	integer	
Date	Calendar date and clock time	character	
JULIAN	Julian day	character	
WEEK	Week number	character	
StartTime	Clock time where data were collected	character	
Sunrise_Time	Sunrise time at a specific date	character	
QC	?	character	
JDY_GSS	Julian day relative to the growing season	character	
NUM_DAYS_GSS	?	character	
MIN_SUN	?	character	
Version	?	character	
Sunrise	?	character	
obs	?	character	

2.5 Point Count Table

Point counts table represents the actual observations recorded at each point counts. An observation is defined by a species code and an abundance. Standard OAU codes are used for species. In addition, each observation is assigned a distance class (Table 11), a time interval class (table 12), and a behavioural code (Table 13). These codes and classes are described in detail below.

Table 10: PointCount Table Structure

Field Name	Field Description	Data Type	Code Description
PCODE	Project code. Allow to link data sharing agreement to stations	character	Section 2.2.1, Table 2
SS	Sampling site identifier. Concatenation of PCODE:SITE:STN	character	
PKEY	Individual point count identifier. Concatenation of SS:YY:ROUND	character	
Duration	Time interval during which the observation was made	character	Section 2.5.2, Table 12
Distance	Distance interval during which the observation was made	character	Section 2.5.1, Table 11
Species	Bird species code according the the standard of the American Ornithological Union	character	
oldspp	?	character	
Abund	Number of birds counted	character	
BEH	Recorded behaviour by which the individual bird or birds were detected	character	Section 2.5.3, Table 13
newid	?	character	

2.5.1 Distance

Distance repesents the distance interval class during which the observation was recorded. Among projects, distance range varies between 0 to unlimited split in various interval class specific to each project. Please refer to DISTMETH table (Table 4) to know more about the distance protocol used by the project.

Table 11: Distance Code Description

Distance Code	Sort	Description
1	1	0-50m
2	2	50-100m
3	3	>100m
4	4	cc edge
5	5	in forest away from clearcut
6	6	»100m (likely diff habitat)
7	7	unlimited distance
8	8	0-100m
9	9	unk
10	10	0-25m
11	11	25-50m
12	12	50-75m
13	13	75-100m
14	14	>75m
15	15	>50m
16	16	100-125m
17	17	125-150m
18	18	0-60m

Table 11: Distance Code Description *(continued)*

Distance Code	Sort	Description
19	19	>60m
20	20	>150m
21	21	100-150m
22	22	100-175m
23	23	0-30m
24	24	30-50m
25	25	0-75m
26	26	0-80m
27	27	0-10m
28	28	10-20m
29	29	20-30m
30	30	30-40m
31	31	40-50m
32	32	50-60m
33	33	60-70m
34	34	70-80m
35	35	80-90m
36	36	90-100m
37	37	100-125m
38	38	10-25m
39	39	recorder

2.5.2 Duration

Duration repesents the period interval during which the observation was recorded. As for distance, duration varies greatly among projects starting with a duration range starting from 3 min to more than 10 minutes count survey. Range are splitted in various interval class specific to each project. Please refer to DURMETH table (Table 3) to know more about duration protocol used by the project.

Table 12: Period Code Description

Duration Code	Description	Sort
1	0-5	1
2	5-10	2
3	before or after	3
4	0-10	4
5	0-3	5
6	3-10	6
7	3-5	7
8	unk	8
9	10-15	9
10	15-20	10
11	0-20	11
12	5-8	12
13	8-10	13
14	3-7	14
15	7-10	15
16	0-3.33	16
17	3.33-6.66	17
18	6.66	18
20	0-1	20
21	1-2	21
22	2-3	22
23	3-4	23
24	4-5	24
25	5-6	25

Table 12: Period Code Description *(continued)*

Duration Code	Description	Sort
26	6-7	26
27	7-8	27
28	8-9	28
29	9-10	29
30	0-2	30
23a	3-4	31
24a	4-5	32
25a	5-6	33
26a	6-7	34
27a	7-8	35
28a	8-9	36
29a	9-10	37

2.5.3 Behaviour

Behaviour indicates if the detection was visual, auditory, or both. It also includes flyovers and flythroughs separately, record sex, age, activity/behaviour data.

Table 13: Behaviour Code Description

Duration Code	Description	Sort
1	0-5	1
2	5-10	2
3	before or after	3
4	0-10	4
5	0-3	5
6	3-10	6
7	3-5	7
8	unk	8
9	10-15	9
10	15-20	10
11	0-20	11
12	5-8	12
13	8-10	13
14	3-7	14
15	7-10	15
16	0-3.33	16
17	3.33-6.66	17
18	6.66	18
20	0-1	20
21	1-2	21
22	2-3	22
23	3-4	23
24	4-5	24
25	5-6	25
26	6-7	26
27	7-8	27
28	8-9	28
29	9-10	29
30	0-2	30
23a	3-4	31
24a	4-5	32
25a	5-6	33
26a	6-7	34
27a	7-8	35
28a	8-9	36
29a	9-10	37