APPENDIX: OVERVIEW OF 2009-2012 VMS ANALYTICAL DATASETS

Murali Mood and Michael Palmer, NEFSC Population Dynamics

Rationale

Beginning in 2009, the Vessel Monitoring System (VMS) data available at the Northeast Fisheries Science Center (NEFSC) came from the HQ-based VMS system. Previously, the dataset available for NEFSC staff contained basic information on VMS activity declaration and vessel name and hull identifier. The content of these data fields both in the historical dataset as well as in the current data set is uncertain and in cases, questionable. For this reason when the data pull for the 2009-present data was initially constructed, these fields were not included. This initiative reflects an attempt to construct a clean VMS analytical set that contains as accurate information as is available in the core data sources (VMS, Allocation Management System [AMS], permits, Vessel Trip Report [VTR], and Observer Database [OBDBS]). Currently these products are available from 2009 onward. Subsequent work will explore extending this time series back in time as supported by the underlying data sets.

Table/view Summary

T_TRIP_OBS_VMS_VTR_yyyy (table): Base table that contains the full matched set of VMS positions, AMS activity declarations, VTR trips and observer trips.
Matching procedures (Appendix Figure 1):

1. VTR data (from VESLOG series) matched using outer join on permit number or hull identifier and trip midpoint. The VESLOG trip midpoint is matched between the OBDBS sail and landing dates. **note VESLOG.TRIPID has been re-coded as DOCID.*
    - Outer join condition: ((obtrp.hullnum1 = veslog.hullnum) OR (obtrp.permit1 = veslog.permit)) AND (veslog.datesail + (veslog.datelnd1 - veslog.datesail) / 2) BETWEEN obtrp.datesail AND obtrp.dateland

2. The VTR/Observer data product was matched to AMS (TRP) data using an outer join on permit number **or** hull identifier and the trip midpoint. The VESLOG trip midpoint is matched between AMS sail and landing dates.
    - Outer join condition: trp.permit_nbr = veslog.permit AND (veslog.date_sail + (veslog.date_land - veslog.date_sail) / 2) BETWEEN trp. date_sail AND trp.date_land
    - Join to the observer data uses the permit number preferentially, if available, then hull identifier.

3. VMS positions were joined to activity declarations using permit number and trip sail/end dates (i.e., pull all VMS positions within the AMS trip/sail dates). **VMS positions were adjusted from GMT to local time prior to the join.**
        i. Outer join condition: T.permit = VMS.vessel_permit AND vms.pos_sent_date BETWEEN T.date_sail AND T.date_land. Note: “T” here is data from step 2 (VTR/Observer join to AMS data).

ii. Since the interest is only in AMS trips with VMS positional info, trips with AMS.TRIP_SOURCE='VMS'.
iii. A trip is included in a calendar year based on the date of landing.
iv. Occassionally, the QA/QC team at RO invalidates some trips by making the activity codes as 'XXX-XXX-XXXXXX'. These trips are either short duration trips, which can be stitched with the actual trip preceding or succeeding this or just flyers. Most of these are scallop trips. These trips are excluded from the AMS data.

V_yyyy_TRIP_MATCH_VMS (view): This is the base analytical view. It contains all VMS positions that could be matched to an AMS trip. It also includes the VTR DOCID (from VESLOG) and OBDBS.LINK1 where a match exists. The match between AMS and VMS is an inner join, the match between AMS/VMS and VTR and OBDBS is a left join (includes only VTR and OBDBS trips than can be matched to the AMS/VMS set). See matching rules outlined above. Complete documentation of the fields contained in this view is presented in Appendix Table 1.

V_yyyy_TRIP_NOVMS (view): These are AMS trips that have no corresponding VMS position information.

V_yyyy_VMS_NOTRIP (view): These are VMS positions that have no matching AMS trip information.

Duplicate Matches

Data quality reviews on the 2011 data revealed the presence of approximately 500 trips where multiple matches existed across databases. For example, single AMS trips would match to multiple VTR trips if an AMS trip had prolonged sail dates (e.g., tripid 557451 has date_sail of 12/10/2010 and date_land of 2/17/2011). This trip would match to many DOCID/LINK1 values. The matched trips have been excluded from the analytical data set and isolated into individual views depending on the type of duplication. There are three categories of duplication which are summarized below:

MV_YYYY_DUPES_DAS_ID: *AMS matches to multiple OBDBS, VTR.*
MV_YYYY_DUPES_DOCID: *VTR matches to multiple OBDBS, AMS.*
MV_YYYY_DUPES_LINK1: *OBDBS matches to multiple VTR, AMS.*

These duplicate matches should be periodically evaluated and forwarded to the appropriate contacts so that data issues can be addressed in the base data sets (AMS, VTR or OBDBS). By querying the primary key in these views back with the T_TRIP_OBS_VMS_VTR_yyyy table, the matched trips from the corresponding database can be determined and investigated.

Calculated Speed and Course

The base VMS data contains both speed and course information. Previous experience with VMS data has found these values to be occasionally unreliable when compared to speed and course estimates calculated from the previous VMS position. Most notably, the VMS data

contain course values of 0 in situations where there is no movement between subsequent positions and course cannot be defined.

An analysis was conducted to compare the speed and course data to the corresponding calculated values. Speeds were calculated based on the Haversine distance equation (Sinnott 1984) from the previous vessel position fields in the VMS data (PREV_LAT_GIS, PREV_LON_GIS, PREV_POS_SENT_DATE). The Haversine distance equation utilizes an approximation of earth curvature that should be sufficient at the scales associated with VMS data.

Results of the speed and course comparison are shown in Appendix Figures 2 and 3, respectively. In general, the data and calculated values agree, but there are notable 'runs' in the data values compared to the calculate values. For example, the calculated speed values show much greater variability than the VMS data values at 0, 3, 6, 9, 12, 15, and 18 knots. **The regularity of the 'run' increments suggests that there is some rounding of the VMS data values and argues for use of the calculated values over the values reported in the base VMS data.**

The calculated speed and course values make use of the previous vessel position fields in the VMS data (PREV_LAT_GIS, PREV_LON_GIS, PREV_POS_SENT_DATE). Since these data would be used as the basis of the calculated speed and course fields, a quick check was conducted to evaluate the accuracy of these data compared to calculated previous positional points, which were calculated using a lag procedure. Once the previous position was calculated, speed and course were calculated. These calculated speed and course values were then compared to the corresponding metrics calculated from the previous vessel position fields in the VMS data (PREV_LAT_GIS, PREV_LON_GIS, PREV_POS_SENT_DATE). The results of this comparison show that there are no differences in the speed or course values calculated from either approach (Appendix Figures 4 and 5). **Because there are no differences, the calculated speed and course data contained in the analytical set will be based on the previous vessel position fields in the VMS data (PREV_LAT_GIS, PREV_LON_GIS, PREV_POS_SENT_DATE).**

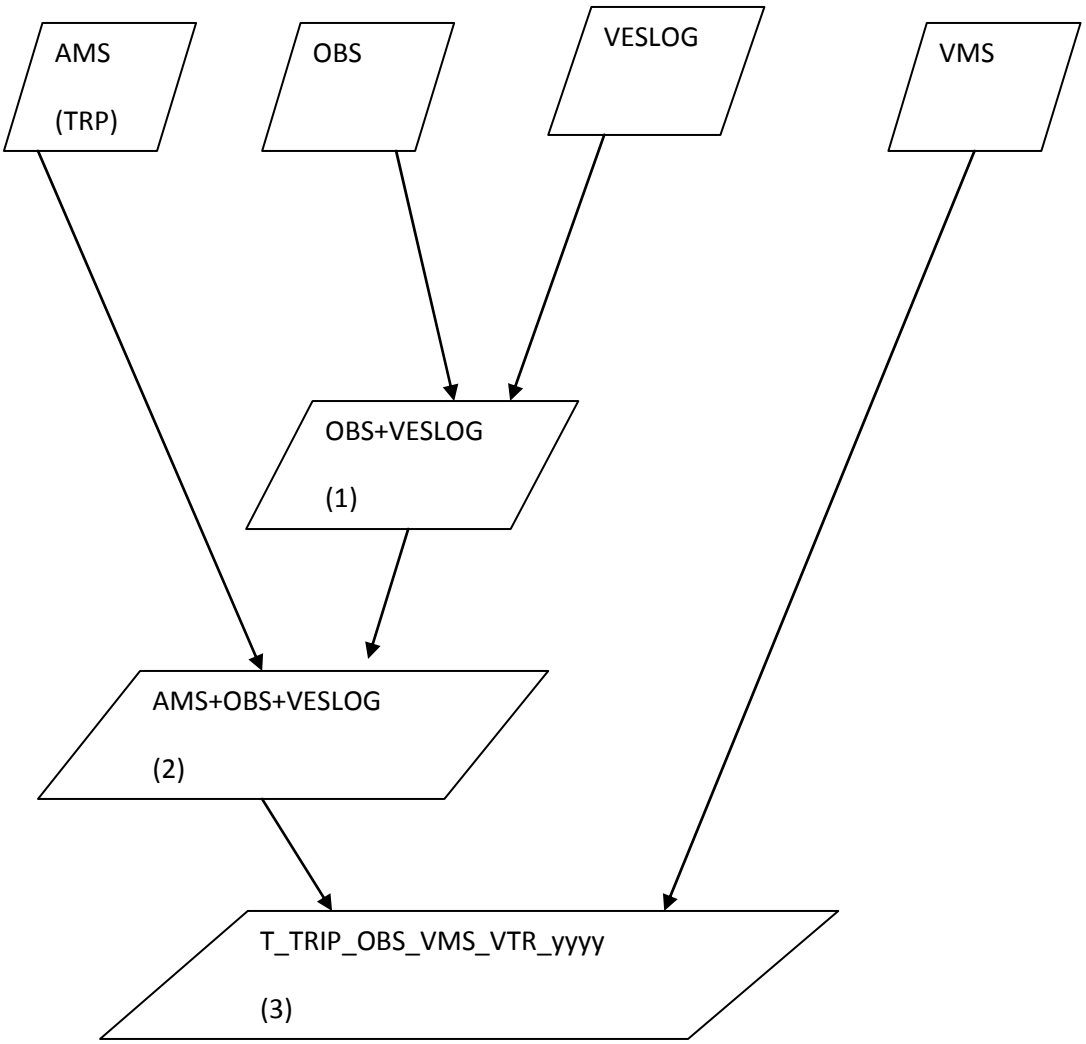
It should be noted that in the 2011 data used for development of the analytical data set 15 records were missing previous position and datetime info (i.e., null). For these records the calculated speed and course cannot be determined. The represents a very small percentage of the total VMS positions available (<0.1 %) and is thus not considered to be problematic.

REFERENCE CITED

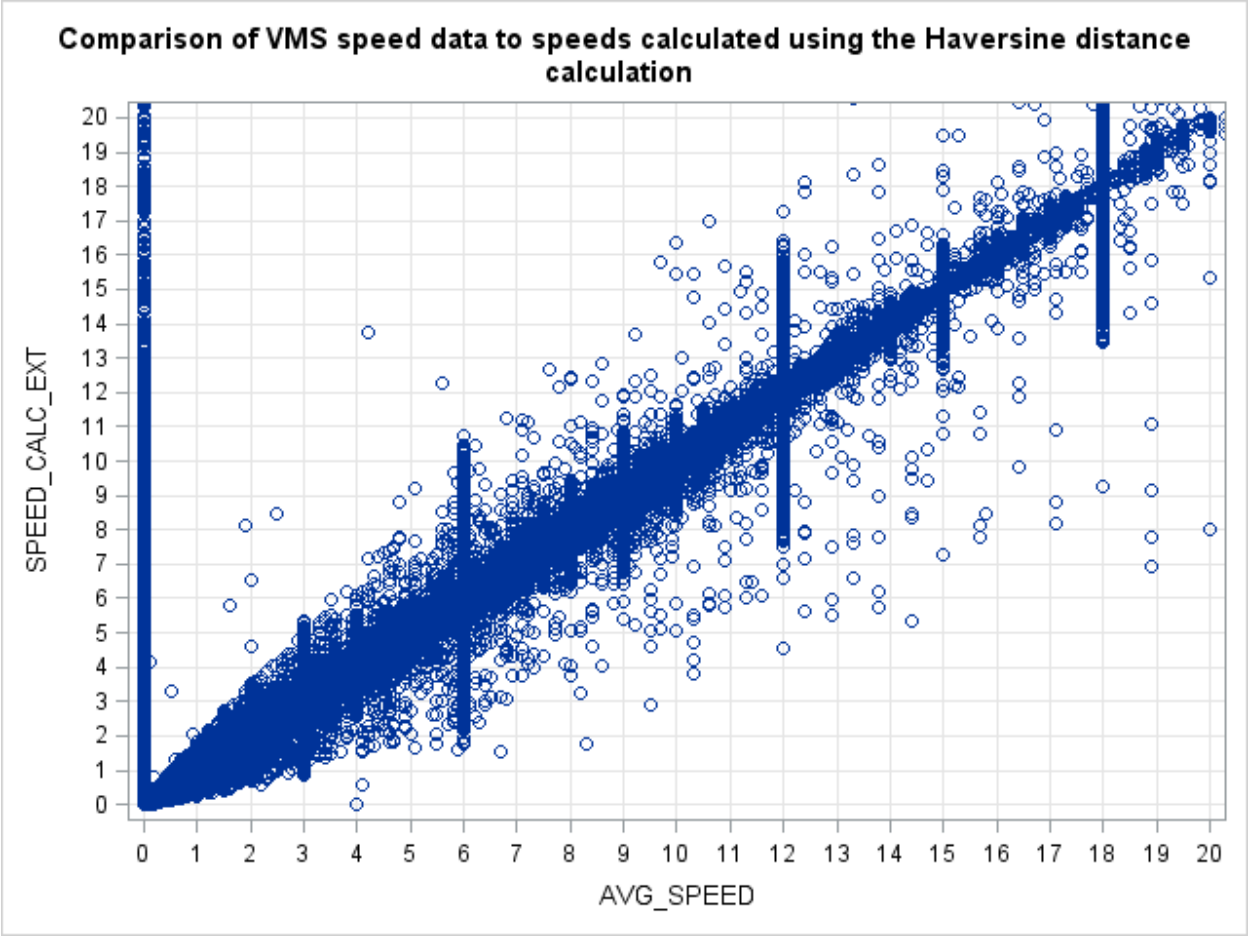
Sinnott, R.W., 1984. Virtues of the haversine. Sky Telescope 68, 159.

Appendix Table 1. Description of the fields contained in the base analytical tables, V_yyyy_TRIP_MATCH_VMS.

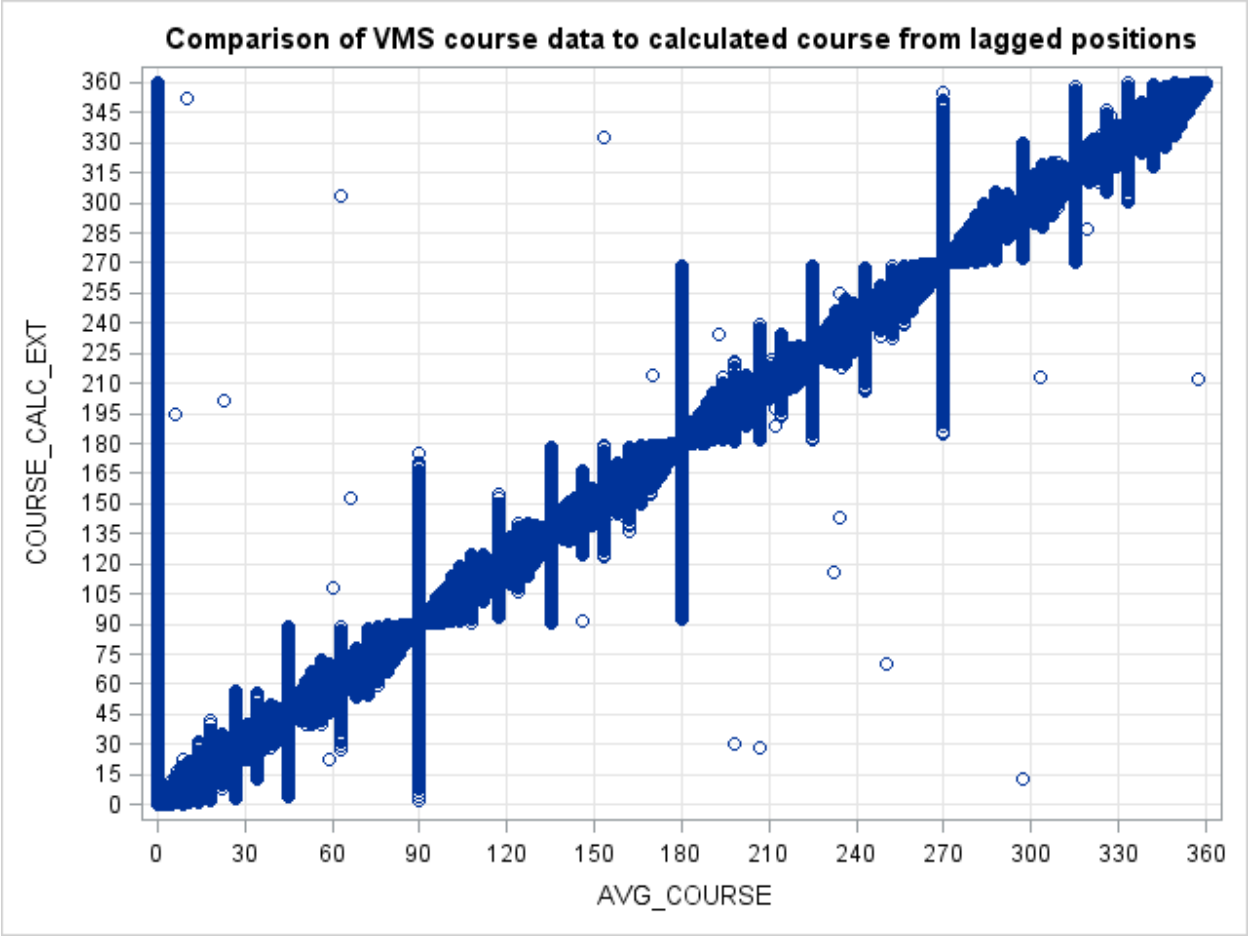
Field	Data Type	Description	Source
PERMIT	Number	Vessel's federal permit number	NOVA.VMS.VMSyyyy.VESSEL_PERMIT
VESSEL_HULL_ID	Varchar2 (10)	Official vessel registration number painted on the hull. US Coast Guard, state or tribal registration number. Unkown=blank/null.	Derived from the SOLE.PERMIT database based on the permit, date_sail and activity code from the trip table. This method has limitations such as not being able to get the FMP from a DOF trip declaration. In these cases the value is set to a null value.
TRIP_ID	Number (10)	This is the ID field in the VMS dataset. Since this field is not contained in the	NOVA.AMS.TRIP.DAS_ID
DAS_ID	Number (10)	This is the unique identifier of the SOLE.AMS.TRIP table.	NOVA.AMS.TRIP.TRIP_ID
ACTIVITY_CODE	Varchar2 (20)	The VMS activity code declared by the vessel.	NOVA.AMS.TRIP.ACTIVITY_CODE
DATE_SAIL	Date	The date (and time) the trip sailed by ams definition (typically the date the vessel crosses the VMS demarcation line).	NOVA.AMS.TRIP.DATE_SAIL
DATE_LAND	Date	The date (and time) the trip ended by ams definition (typically the date the vessel crosses the VMS demarcation line).	NOVA.AMS.TRIP.DATE_LAND
POS_SENT_DATE	Date	The date (and time) the positional information was sent from the VMS unit.	NOVA.VMS.VMSyyyy.POS_SENT_DATE
LAT_GIS	Varchar2 (11)	The vessels latitude at the time the position information was sent from the VMS unit (decminal degrees).	NOVA.VMS.VMSyyyy.LAT_GIS
LON_GIS	Varchar2 (11)	The vessels longitude at the time the position information was sent from the VMS unit (decminal degrees).	NOVA.VMS.VMSyyyy.LON_GIS
PREV_POS_SENT_DATE	Date	The previous date (and time) the vessel transmitted positional information from the VMS.	NOVA.VMS.VMSyyyy.PREV_POS_SENT_DATE
PREV_LAT_GIS	Varchar2 (11)	The vessels latitude at theprevious position (decminal degrees).	NOVA.VMS.VMSyyyy.PREV_LAT_GIS
PREV_LON_GIS	Varchar2 (11)	The vessels longitude at the previous position (decminal degrees).	NOVA.VMS.VMSyyyy.PREV_LON_GIS
VMS_AVG_COURSE	Number (3)	**Vessel course (degrees). Use with caution!**	NOVA.VMS.VMSyyyy.AVG_COURSE
VMS_AVG_SPEED	Number (4,1)	**Vessel speed (knots). Use with caution!**	NOVA.VMS.VMSyyyy.AVG_SPEED
CALC_AVG_SPEED	Number	Calculated vessel speed (knots) from positional and previous positional information.	
CALC_AVG_COURSE	Number	Calculated vessel course (degrees) from positional and previous positional information.	
HULLNUM	Varchar2 (10)	Hull identifier from matched VTR or observer trips. When the trip can be matched to both VTR and observer data, the hull identifier is taken preferentially from the VTR data.	SOLE.VTR.VESLOGyyyyT.HULLNUM or NOVA.OBDBS.OBTRP.HULLNUM1
DOCID	Number	The primary key in the VTR VESLOGyyyyT table	SOLE.VTR.VESLOGyyyyT.TRIPID
LINK1	Varchar2 (15)	The primary key in the OBTRP table	NOVA.OBDBS.OBTRP.LINK1



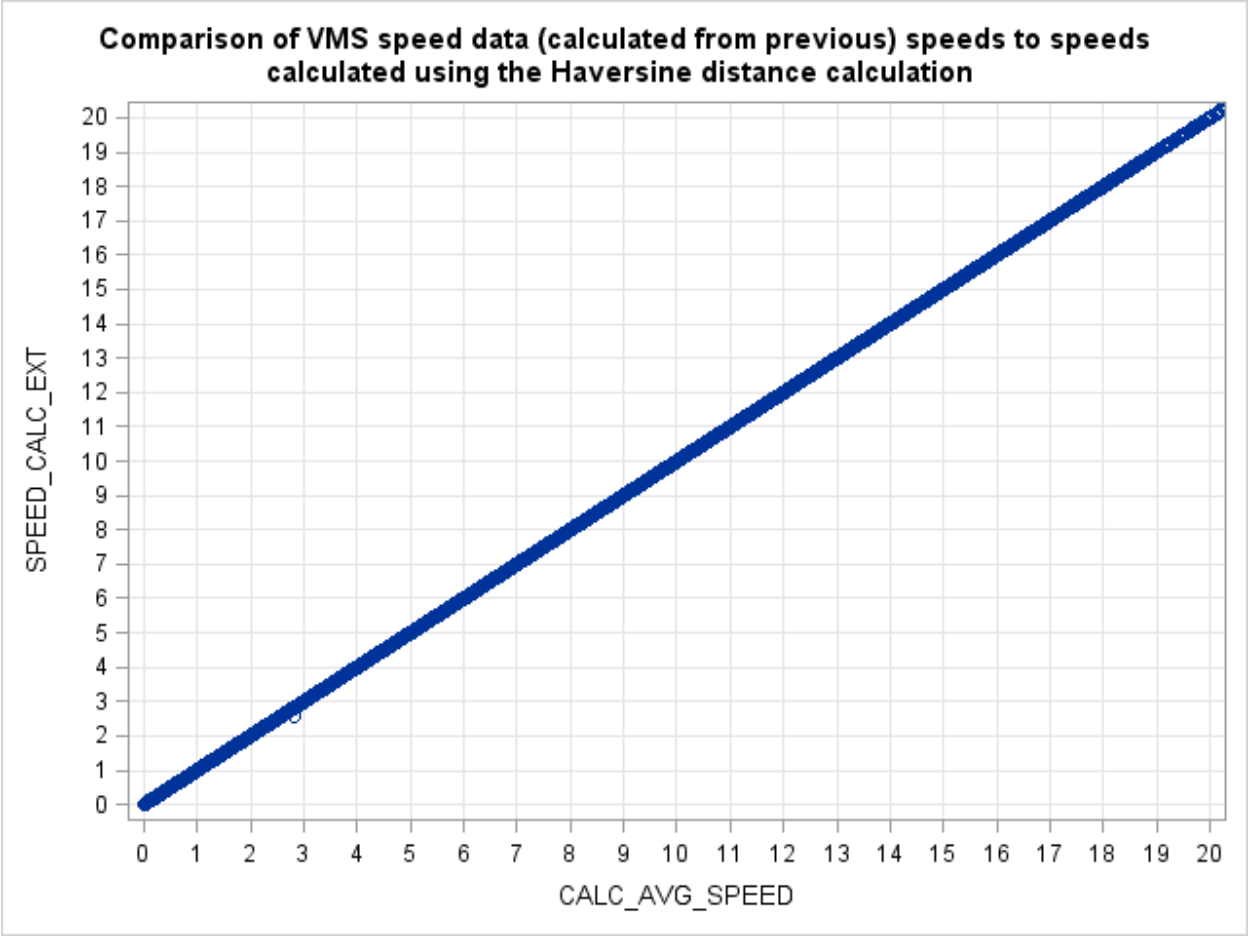
Appendix Figure 1. Schematic diagram of the data set joins used to generate the final T_TRIP_OBS_VMS_VTR_yyyy product.



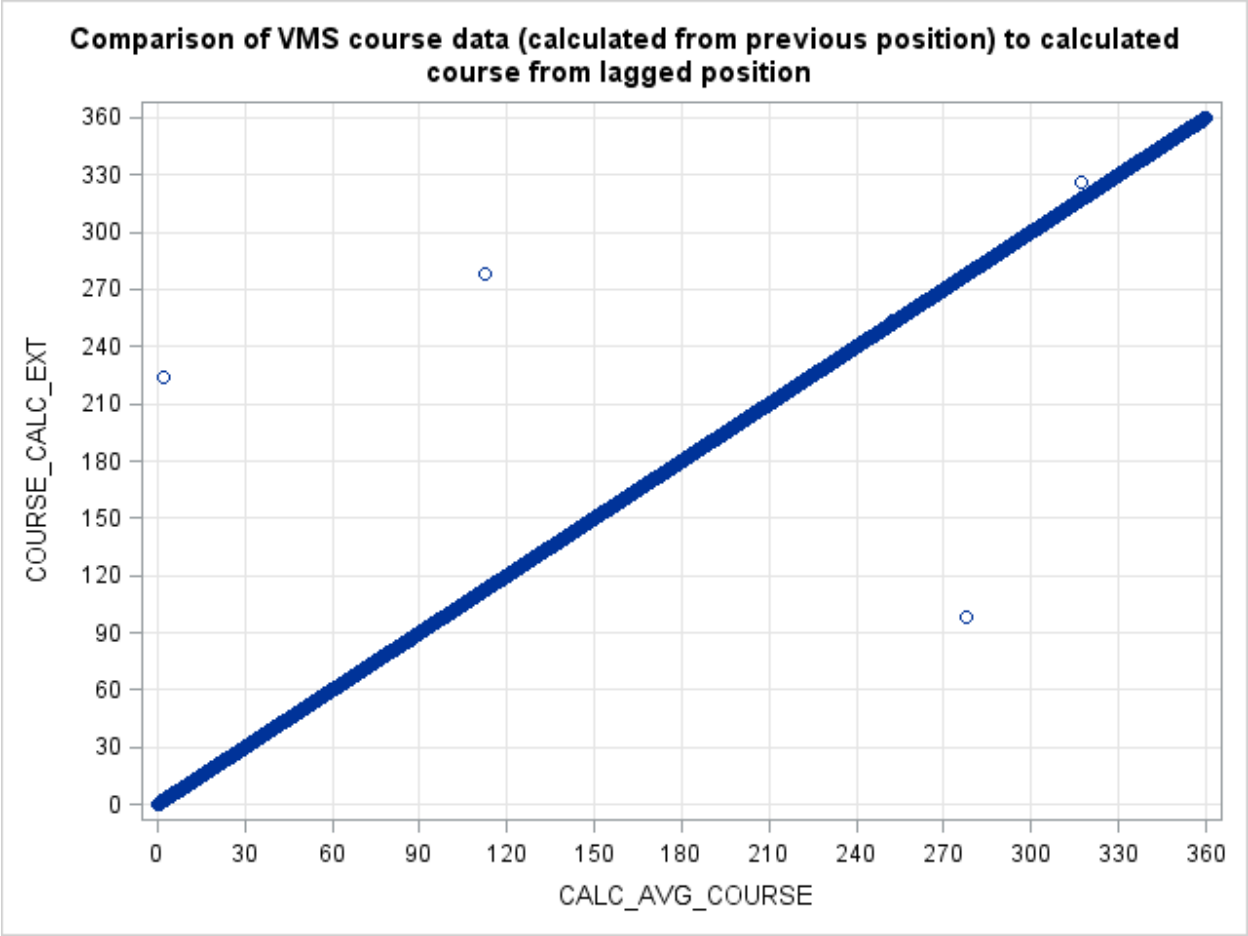
Appendix Figure 2. Comparison of the calculated average speed (CALC_AVG_SPEED) in knots to the speed value contained in the VMS data (VMS_AVG_SPEED).



Appendix Figure 3. Comparison of the calculated average course (CALC_AVG_COURSE) in degrees to the course value contained in the VMS data (VMS_AVG_COURSE).



Appendix Figure 4. Comparison of the average speed calculated from a determined previous position (SPEED_CALC_EXT) to the average speed calculated from the previous position reported in the VMS data (CALC_AVG_SPEED).



Appendix Figure 5. Comparison of the average course calculated from a determined previous position (COURSE_CALC_EXT) to the average course calculated from the previous position reported in the VMS data (CALC_AVG_COURSE).