SHIPS Predictor Files

Last Updated: 2 Oct 2023

Modifications for 2023: 2022 cases added for AL, CP and EP.

Important note on file errors: The AL, CP and EP files were first updated on 23 June 2023. However, a perturbation was added to the tracks in an experimental version being used to evaluate the impact of overfitting in the SHIPS model. The perturbation was zero in the t=0 h positions in the file so the t=0 h predictor values are correct, but the track perturbations were included starting at 6 hr and increased with time out to 168 h. The experimental lsdiag.dat versions were accidentally updated on 23 June 2023. The corrected lsdiag.dat files were uploaded at 2:35 PM UTC on 03 Oct 2023. This error was not made in the WP, IO or SH files.

Modifications for 2022: 2021 cases added for AL, CP and EP.

Modifications for 2021: 2020 cases added for AL, CP and EP. 2019 cases updated with final best tracks. 7-day files now available.

Modifications for 2020: 2018 and 2019 cases added for AL, CP, and EP.
Preliminary Best Tracks were used for 2019 cases. Final Best Tracks were used for 2018 cases.

Modifications for 2018: 2017 cases added for AL, EP, and CP
Daily Reynolds SST variables added (DSST, DSTA) Time dependent model TPW variables added
Subsurface ocean variables added from NCODA analysis
PHCN variable removed (proxy for ocean heat content)

Modifications for 2017: 2016 cases added for AL, EP, CP (preliminary best track data)
Missing EP cases from 1983 restored
More IR satellite variables added

Modifications for 2016: 2015 cases added for AL, EP, and CP
TPW variables from GFS analysis added (MTPW)
Environmental storm-relative helicity (HE07, HE05)
Large-scale vertical velocity from GFS added (O500, O700)
"Ocean Age" variables added (OAGE, NAGE) Extra time added to IR variables (IRM1)

Modifications for 2015: 2014 cases added for AL, EP, and CP.
Climate Forecast System Reanalysis (CFSR) fields used for 1982- 1999 cases instead of old NCAR/NCEP Reanalysis.
Three new variables added (G150, G200, and G250)

Modifications for 2014: 2013 cases added for AL, EP, and CP

Modifications for 2013: 2012 cases added
New climatological variables added (CD20, CD26, COHC).
East and Central Pacific data in separate files.

Starting point for Atlantic, East and Central Pacific files is the new NHC HURDAT2 file. All unnamed cyclones prior to 1989 were eliminated.

Modifications for 2012: 2011 cases added
New fields added to the PSLV line
West Pacific file added

Modifications for 2011: 2010 cases added
Six new variables added (V000, V850, V500, V300, TGRD, TADV)

Modifications for 2010: 2009 cases added
Modification to the header line (min pressure and ATCF ID added)
TYPE line added (indicates storm type)
HIST line added (storm age variable)

Modifications for 2009: 2008 cases added
Older climatological ocean variables deleted (D20C, D26C, HCON)
New climatological heat content variable (PHCN)
GOES predictors extended back to 1983 (from 1995 in previous ver.)
Ocean heat content (RHCN) now available for east Pacific (to 1996)
New experimental predictors added (U20C, V20C, VVAC)

Modifications for 2008: 2007 cases added, and two new experimental predictors added for testing (VVAV and VMFX)

Modifications for 2007: Several new synoptic variables related to storm vortex removal

Modifications for 2006: 2 new synoptic variables (T250, T150)
6 new GOES IR predictors

Modifications for 2005: 12-hour records/data replaced by 6-hour records/data
4 new synoptic variables (T000, Z000, R000)
GOES IR data predictors at t=-3 hr (IRM3)

Currently available files:

- lsdiaga_1982_2023_sat_ts.dat – Atlantic data with predictors from CFSR re-analysis Fields 1982-2000 and operational GFS analyses 2001-present with satellite variables when available (SAT) and cases are either tropical or subtropical (TS)

- lsdiage_1982_2023_sat_ts.dat Same as the above file for the East Pacific

- lsdiagc_1982_2023_sat_ts.dat Same as the above file for the Central Pacific

- lsdiagw_1990_2020_rean_sat_nbc_ts.dat. Same as the above file for the Western North Pacific, but with CFSR input from 1990-2004 and GFS operational analyses from 2005-present.

- lsdiagi_1990_2020_rean_sat_nbc_ts.dat. Same as the above file for the N. Indian Ocean

- lsdiags_1998_2020_rean_sat_nbc_ts.dat. Same as the above file for the S. Hemisphere

The SHIPS predictor file is divided into a set of records for each storm case. Each line of the file ends with a line descriptor, and each set of records starts with HEAD and ends with LAST.

The current files include cases at 6-hour intervals. Not all predictors are available for all years (see note below).

Each line of the file except HEAD, HIST, IRXX, IR00, IRM1, IRM3, PSLV, MTPW and LAST contain predictors at 6-hour intervals (-12, -6, 0, 6, 12 ... 120 hr) relative to the time and date of the particular case. The new 7-day files extend to 168 hr. If the value of a predictor is not available, the field contains 9999. All of the fields are relative to the storm center, where the storm center is determined from the NHC Best Track, except where specified. All of the atmospheric predictors are from NCEP global model analyses (either re-analysis or operational). The predictors in the file contain the following:

HEAD: Header line (1st four letters of storm name, 2-digit year, month, day, and UTC time, maximum winds (kt), lat, lon, minimum sea level pressure (hPa), and ATCF ID number (e.g., AL011982) at t=0 of current case.
TIME: Time relative to current case (hr)
VMAX: Maximum surface wind (kt)
MSLP: Minimum Sea Level Pressure (hPa)
TYPE: Storm type (0=wave, remnant low, dissipating low, 1=tropical, 2=subtropical, 3=extra-tropical). Note that the SHIPS variables are set to missing for all cases except type=1 or 2, since these are not included in the SHIPS developmental sample for estimating the model coefficients.
HIST: Storm history variable. The no. of 6 hour periods the storm max wind has been above 20, 25, ,…, 120 kt.
DELV: Intensity change (kt) -12 to 0, -6 to 0, 0 to 0, 0 to 6, ... 0 to 120 hr. If the storm crosses a major land mass during the time interval, value set to 9999
INCV: Intensity change (kt) -18 to -12, -12 to -6, ... 114 to 120 hr. INCV is set to 9999 similar to DELV for land cases.
LAT: Storm latitude (deg N *10) vs time
LON: Storm longitude (deg W *10) vs time
CSST: Climatological SST (deg C * 10) vs time
CD20: Climatological depth (m) of 20 deg C isotherm from 2005-2010 NCODA analyses
CD26: Same as CD20 for the 26 deg C isotherm
COHC: Same as above for ocean heat content (kJ/cm^2)
DTL: Distance to nearest major land mass (km) vs time
OAGE: Ocean Age (hr*10), which is the amount of time the area within 100 km of the storm center has been occupied by the storm along its track up to this point in time
NAGE: Same as OAGE, but multiplied by the maximum wind/100kt. If the max wind was a constant 100 kt over its past history, NAGE=OAGE.
RSST: Reynolds SST (deg C*10) vs time. Number after SST label is the age in days of the SST analysis used to estimate RSST from the weekly Reynolds SST analyses.
DSST: Same as RSST, but from daily Reynolds SST analyses
DSTA: Same as DSST, but spatially averaged over 5 points (storm center, + 50 km N, E, S and W of center)

U200: 200 hPa zonal wind (kt *10) vs time (r=200-800 km)
U20C: Same as U200 but for r=0-500 km)
V20C: Save as U20C, but for the v component of the wind
E000: 1000 hPa theta_e (r=200-800 km) vs. time (deg K*10)
EPOS: The average theta e difference between a parcel lifted from the surface and its environment (200-800 km average) versus time (deg C * 10). Only positive differences are included in the average. **ENEG:** Same as EPOS, but only negative differences are included. The minus sign is not included.
EPSS: Same as EPOS, but the parcel theta_e is compared with the saturated theta_e of the environment
ENSS: Same as ENEG, but the parcel theta_e is compared with the saturated theta_e of the environment
RHLO: 850-700 hPa relative humidity (%) vs time (200-800 km)
RHMD: Same as RHLO for 700-500 hPa
RHHI: Same as RHLO for 500-300 hPa
PSLV: Pressure of the center of mass (hPa) of the layer where storm motion best matches environmental flow (t=0 only) and the information used to calculate the steering layer pressure. All fields are valid at t=0, and those in the t=6 to t=102 columns include the following:

t= 6 column: The observed zonal storm motion component (m/s *10) t= 12 column: The observed meridional storm motion component (m/s *10) t=18, t=24 columns: Same as t=6,12 hr columns but for the 1000 to 100 hPa mass weighted deep layer environmental wind (m/s *10)

t=30, t=36 columns: Same as t=6,12 columns but for the optimally weighted deep layer mean flow (m/s *10)

t=42 column: The parameter alpha that controls the constraint on the weights from being not too "far" from the deep layer mean weights (non-dimensional, *100)

t=48 to t=102 columns: The optimal vertical weights for p=100, 150, 200, 250, 300, 400, 500, 700, 850 and 1000 hPa. (non-dimensional *1000)

Z850: 850 hPa vorticity (sec-1 * 10**7) vs time (r=0-1000 km)
D200: Same as above for 200 hPa divergence
REFC: Relative eddy momentum flux convergence (m/sec/day, 100-600 km avg) vs time
PEFC: Planetary eddy momentum flux convergence (m/sec/day, 100-600 km avg) vs time
T000: 1000 hPa temperature (dec C* 10) (200-800 km average)
R000: 1000 hPa relative humidity (200-800 km average)
Z000: 1000 hPa height deviation (m) from the U.S. standard atmosphere
TLAT: Latitude of 850 hPa vortex center in NCEP analysis (deg N*10)
TLON: Longitude of 850 hPa vortex center in NCEP analysis (deg W*10)
TWAC: 0-600 km average symmetric tangential wind at 850 hPa from NCEP analysis (m/sec *10)
TWXC: Maximum 850 hPa symmetric tangential wind at 850 hPa from NCEP analysis (m/sec *10)
G150: Temperature perturbation at 150 hPa due to the symmetric vortex calculated from the gradient thermal wind. Averaged from r=200 to 800 km centered on input lat/lon (not always the model/analysis vortex position). (deg C*10)
G200: Same as G150 at 200 hPa
G250: Same as G150 at 250 hPa
V000: The tangential wind (m/sec *10) azimuthally averaged at r=500 km from (TLAT,TLON) If TLAT,TLON are not available, (LAT,LON) are used.
V850: Same as V000 at 850 hPa

V500: Same as V000 at 500 hPa
V300: Same as V000 at 300 hPa
TGRD: The magnitude of the temperature gradient between 850 and 700 hPa averaged from 0 to 500 km estimated from the geostrophic thermal wind (deg C per m*10^7)
TADV: The temperature advection between 850 and 700 hPa averaged from 0 to 500 km Estimated from the geostrophic thermal wind (deg per sec*10^6)
PENC: Azimuthally averaged surface pressure at outer edge of vortex ((hPa-1000)*10)
SHDC: Same as SHRD but with vortex removed and averaged from 0-500 km relative to 850 hPa vortex center
SDDC: Heading (deg) of above shear vector. Westerly shear has a value of 90 deg.
SHGC: Same as SHRG but with vortex removed and averaged from 0-500 km relative to 850 hPa vortex center
DIVC: Same as D200, but centered at 850 hPa vortex location
T150: 200 to 800 km area average 150 hPa temperature (deg C *10) versus time
T200: Same as above for 200 hPa temperature (deg C *10)
T250: Same as above for 250 hPa temperature (deg C *10)
SHRD: 850-200 hPa shear magnitude (kt *10) vs time (200-800 km)
SHTD: Heading (deg) of above shear vector. Westerly shear has a value of 90 deg.
SHRS: 850-500 hPa shear magnitude (kt *10) vs time
SHTS: Heading of above shear vector
SHRG: Generalized 850-200 hPa shear magnitude (kt *10) vs time (takes into account all levels from 1000 to 100 hPa
PENV: 200 to 800 km average surface pressure ((hPa-1000)*10)
VMPI: Maximum potential intensity from Kerry Emanuel equation (kt)
VVAV: Average (0 to 15 km) vertical velocity (m/s * 100) of a parcel lifted from the surface where entrainment, the ice phase and the condensate weight are accounted for. Note: Moisture and temperature biases between the operational and reanalysis files make this variable inconsistent in the 2001-2007 sample, compared 2000 and before.
VMFX: Same as VVAV, but a density weighted vertical average.
VVAC: Same as VVAV but with soundings from 0-500 km with GFS vortex removed
HE07: Storm motion relative helicity (m^2/s^2)*10 for p=1000 to 700 hPa, r=200 to 800 km
HE05: Same as HE05, for P=1000 to 500 hPa
O500: Pressure vertical velocity (hPa/day) at 500 hPa, averaged from r=0 to 1000 km
O700: Same as O500 at 700 hPa
CFLX: Dry air predictor based on the difference in surface moisture flux between air with the observed (GFS) RH value, and with RH of air mixed from 500 hPa to the surface.
MTPW: Total Precipitable Water (TPW) predictors at t=0 from the GFS analysis. The 21 values in this record are as follows:

1) 0-200 km average TPW (mm * 10)
2) 0-200 km TPW standard deviation (mm * 10)
3) 200-400 km average TPW (mm * 10)
4) 200-400 km TPW standard deviation (mm * 10)
5) 400-600 km average TPW (mm * 10)
6) 400-600 km TPW standard deviation (mm * 10)
7) 600-800 km average TPW (mm * 10)
8) 600-800 km TPW standard deviation (mm * 10)
9) 800-1000 km average TPW (mm * 10)
10) 800-1000 km TPW standard deviation (mm * 10)
11) 0-400 km average TPW (mm * 10)
12) 0-400 km TPW standard deviation (mm * 10)
13) 0-600 km average TPW (mm * 10)

14) 0-600 km TPW standard deviation (mm * 10)
15) 0-800 km average TPW (mm * 10)
16) 0-800 km TPW standard deviation (mm * 10)
17) 0-1000 km average TPW (mm * 10)
18) 0-1000 km TPW standard deviation (mm * 10)
19) %TPW less than 45 mm, r=0 to 500 km in 90 deg azimuthal quadrant centered on upshear direction
20) 0-500 km averaged TPW (mm * 10) in 90 deg up-shear quadrant
21) 0-500 km average TPW (mm * 10)

PW01-PW19: Time dependent versions of the 21 TPW variables listed above.

IRXX: Same as IR00 below but generated from other predictors (not satellite data). These should only be used to fill in for missing IR00 if needed.

XDST: Climatological value of the daily Reynolds SST (deg C*10)

XNST-XO20: Climatological values of the NCODA variables listed below (NSST through NO20, except there is no X version of NSTA).

NSST: SST from the NCODA analysis (deg C*10)

NSTA: Same as NSST, but spatially averaged over 5 points (storm center, + 50 km N, E, S and W of center)

NTMX: Max ocean temperature in the NCODA vertical profile (deg C*10)

NDMX: Depth of the max ocean temperature in the profile (m)

NDML: Depth of the mixed layer, defied and the depth where the T is 0.5 colder than at the surface (m). In rare cases, this is the depth where the T is 0.5 warmer than at the surface. In those cases, NDML is negative.

ND30-ND16: Depth of the 30, 28, …, 16 deg C isotherms (m)

NDFR: Depth of the lowest model level in the NCODA analysis (m).

NTFR: Ocean T at the lowest level in the NCODA analysis (deg C*10)

NOHC: Ocean heat content from the NCODA analysis (kJ/cm^2) relative to the 26 C isotherm

NO20: Same as NOHC with respect to the 20 deg C isotherm

IR00: Predictors from GOES data (not time dependent). The 20 values in this record are as follows:

1) Time (hhmm) of the GOES image, relative to this case
2) Average GOES ch 4 brightness temp (deg C *10), r=0-200 km
3) Stan. Dev. of GOES BT (deg C*10), r=0-200 km
4) Same as 2) for r=100-300 km
5) Same as 3) for r=100-300 km
6) Percent area r=50-200 km of GOES ch 4 BT < -10 C
7) Same as 6 for BT < -20 C
8) Same as 6 for BT < -30 C
9) Same as 6 for BT < -40 C
10) Same as 6 for BT < -50 C
11) Same as 6 for BT < -60 C
12) max BT from 0 to 30 km radius (deg C*10)
13) avg BT from 0 to 30 km radius (deg C*10)
14) radius of max BT (km)
15) min BT from 20 to 120 km radius (deg C*10)
16) avg BT from 20 to 120 km radius (deg C*10)
17) radius of min BT (km)
18-20) Variables need for storm size estimation

IRM1: Same as IR00 but at 1.5 hours before initial time

IRM3: Same as IR00 but at three hours before initial time

PC00: Principal components and related variables from IR imagery at t=0

PCM1: Same as PC00 but for 1.5 hours before initial time
PCM3: Same as PC00 but for three hours before initial time
LAST: The last line for this case