

Parallel Implementation of the Ensemble Empirical Mode Decomposition (PEEMD) and Its Application for Earth Science Data Analysis

Bo-Wen Shen¹, Samson Cheung², Yu-Ling Wu³

Frank Li⁴ and David Kao²

¹*San Diego State University, San Diego, CA*

²NASA Ames Research Center, Moffett Field, CA

³ University of Alabama in Huntsville, Huntsville, AL

⁴California Technology Institute, Pasadena CA

bshen@mail.sdsu.edu

Abstract

To efficiently perform multiscale analysis of high-resolution, global, multiple-dimensional data sets, we have deployed the parallel ensemble empirical mode decomposition (PEEMD) package by implementing three-level parallelism into the ensemble Empirical Mode Decomposition (EMD), achieving a scaled performance of 5,000 cores. In this study, we discuss the implementation of the PEEMD and its application for the analysis of Earth science data, including the solution of Lorenz model, an idealized terrain-induced flow and real case Hurricane Sandy (2012), the latter of which is the second costliest hurricane in the US history.

25 *Keywords:* *Parallel Ensemble Empirical Mode Decomposition, Supercomputing*
26 *Computing, Multiscale Analysis, Lorenz model, Nonlinear Chaotic Solutions,*
27 *Tropical Waves, Hurricane.*

28 1 Introduction

29 The core technology of the PEEMD (see List 1 for a complete list of acronyms
30 used in this article and their definitions) is the EMD and ensemble EMD (EEMD)
31 methods, which show remarkable performance in revealing multiscale processes of
32 the data that are non-stationary and nonlinear. The EMD method, which was invented
33 by Dr. Huang and his colleagues in the late 1990s, has been widely applied in many
34 different fields (e.g., Huang et al., 1998). It decomposes one set of observational data
35 into the so-called intrinsic mode functions (IMFs), which represent the oscillatory
36 components of the data. The EEMD was developed to overcome the scale mixing (or
37 mode mixing) problem of the EMD by Wu et al. (2009). By introducing finite
38 amplitude white noises, the EEMD deals with an ensemble of data sets. Each data
39 member, which contains the original observational data and finite amplitude white
40 noise, is decomposed by the EMD to obtain a set of IMFs. The final IMFs are
41 determined as the ensemble average of the multiple sets of IMFs, each of which
42 corresponding to one data member. The EEMD significantly reduces the chance of
43 scale mixing, and it still preserves a filter bank property (e.g., Flandrin et al., 2003;
44 We et al., 2004), i.e., the decomposed mean IMFs staying within the natural filter
45 period windows.

46

47 As a result, the computational resources required for an EEMD run are
48 tremendous, linearly proportional to the number of ensemble trials. Furthermore,
49 applying the one-dimensional (1D) EEMD to analyzing multi-dimensional data or a
50 variety of fields at multiple vertical levels could require substantial computational
51 time. Thus, a parallel version of the EEMD (PEEMD) with a three-level parallelism

52 has been developed to reduce the time required for data decomposition. We briefly
53 introduce the design of the parallelism in the PEEMD and its computational
54 performance in section 2. The scientific performance of the PEEMD in the analysis
55 of Earth science data is discussed in section 3. A summary is given in the final
56 section.

57 2 The EMD, EEMD and PEEMD Methods

58 With the EMD, we can extract the IMFs from the time (or spatial) series of the
59 raw data $X(j)$ and represent it by the first m IMFs and the residual of data (R_m) as
60 follows:

$$X(j) = \sum_{i=1}^m C_i(j) + R_m(j). \quad (1)$$

61 $X(j)$ represents the value of X at $t = j\Delta t$ or $x = j\Delta x$, here Δt and Δx are the temporal
62 and spatial increments, respectively. $C_i(j)$ represents the i^{th} IMF. Next, we introduce
63 the method and processes for EMD/EEMD for calculation of the IMFs and then
64 discuss the multiple-level parallelism in the EEMD. The following notation is
65 adopted:

- 66 • the index “ i ” used to indicate the i^{th} IMF;
- 67 • “ t ” as the independent variable that can be time, longitude, or latitude;
- 68 • $X(t)$ as the original data;
- 69 • $S(t)$ as the data to be shifted out;
- 70 • $C_i(t)$ as the i^{th} IMF;
- 71 • $R(t)$ as the residual of the data; and
- 72 • $U(t)$ ($L(t)$) as the upper (lower) envelope of the given data by connecting
73 maxima (minima) with a cubic spline.

74 Therefore, $R(t) = X(t)$, and $S(t) = R(t)$ as $i=0$, initially. $S(t)$ and $R(t)$ are changed after
75 the beginning of the so-called shifting processes that are summarized as follows:

76 1. $i = 0$ and $R(t) = X(t)$, here $X(t)$ is the original data;
 77 2. Let $S(t) = R(t)$;
 78 3. Identify the extrema (both maxima and minima) of the data $S(t)$;
 79 4. Generate the upper (lower) envelope, $U(t)$ ($L(t)$);
 80 5. Determine the local mean, $m(t) = U(t) + L(t)$, by averaging the envelopes;
 81 6. $S(t) = S(t) - m(t)$, i.e., subtract out the mean from the data;
 82 7. Go to step (3) unless stoppage criteria for an IMF are met;
 83 8. $i = i+1$, $C_i(t) = S(t)$, $R(t) = R(t) - C_i(t)$, and go to step (2).

84 For each IMF, we performed 10 iterations to obtain the responding IMF, i.e. Step 2 to
 85 Step 7 were repeated 10 times.

86

87 Based on the above sifting processes, IMFs possess the following features. The
 88 IMFs are symmetric with respect to the local zero mean and have the same number of
 89 zero crossings and extrema. The IMFs are time (spatial) domain functions that
 90 represent the local variability of the original signal at a particular range of
 91 frequencies (wavelengths). Mathematically, a strict definition of the IMF is as
 92 follows: $U(t) + L(t) = 0$ (Wang et al., 2010), namely its local means are zero.

93 Using the above processes, the IMFs and residual functions are defined at the
 94 same grid points as those for the raw data. The residual, R_m , represents the
 95 differences between the raw data and a sum of the first m IMFs. Similarly, $X(j)$ can
 96 also be represented by the first $m+1$ IMFs and R_{m+1} , as follows:

$$X(j) = \sum_{i=1}^{m+1} C_i(j) + R_{m+1}(j). \quad (2)$$

97 Equations (1) and (2) lead to:

$$R_i(j) = R_{i+1}(j) + C_{i+1}(j).$$

98 Since $C_{i+1}(j)$ is purely oscillatory with respect to $R_{i+1}(j)$, we can consider $R_{i+1}(j)$ as
 99 the local mean of $R_i(j)$. Thus, EMD is a “Reynolds type” decomposition for sifting

100 out (or extracting) periodic components from the data by separating the local mean
101 from the fluctuations using spline fits (Huang et al., 1998).

102

103 For the ensemble EMD, the following steps were performed:

- 104 i. Add a noise series to the targeted data;
- 105 ii. Decompose the data with added noise to IMFs using the EMD;
- 106 iii. Iterate Steps 1 and 2 repeatedly, but with a different noise series each time;
- 107 iv. Obtain the (ensemble) means of the corresponding IMFs of the decompositions as the final result.

109 With the EEMD, the chance of scale mixing is reduced while the dyadic property is
110 preserved. As indicated by Wu et al. (2009), for the mean IMFs, added noises cancel
111 out one another.

112 2.1 Parallel implementation of the EEMD

113

114 In this study, performance results were obtained from the NASA Ames Pleiades
115 supercomputer. The Pleiades supercomputer is an SGI Altix ICE system with a peak
116 performance of 4.97 Pflop/s. With 663.6 Tbytes of total memory, and 185,344 cores,
117 it achieves Linpack performance of 4.089 Pflop/s (as of November 2015). The
118 system contains the following types of Intel Xeon processors: E5-2680v4
119 (Broadwell), E5-2680v3 (Haswell), E5-2680v2 (Ivy Bridge), E5-2670 (Sandy
120 Bridge), and X5670 (Westmere) —to reach different needs and capacities on
121 different NASA projects. Each node (physical blade unit) has 2 sockets, and each
122 socket has an 8-core Sandy Bridge or 10-core Ivy Bridge CPU. Thus, each node has
123 16 Sandy Bridge cores or 20 Ivy Bridge cores. While Sandy Bridge cores (E5-2670)
124 are used in the first two benchmarks, Ivy Bridge cores (E5-2680v2) are used in the
125 third benchmark.

126

127 In the PEEMD for this study, a three-level parallelism is achieved as follows: (1)
128 Decompose data spatial domains into subdomains, (2) Divide ensemble members
129 into different groups, and (3) Split a big loop regarding the shifting processes into
130 several chunks. The first two levels of parallelism use Message Passing Interface
131 (MPI) processes to perform tasks in parallel, while the third level parallelism uses
132 OpenMP multi-threads to perform calculations in parallel. In this section, we discuss
133 three benchmarks to compare the computational performance of the first level
134 parallelism with MPI processes, the fine grained parallelism with OpenMP threads,
135 and the whole three level parallelism with a hybrid of MPI processes and OpenMP
136 threads, respectively. Two data sets used in this study are Extended Reconstructed
137 Sea Surface Temperature (ERSST, Smith et al., 2008; Wu et al., 2009) for the first
138 two benchmarks and an idealized solution of the Mixed Rossby-gravity (MRG, Shen
139 et al., 2010a) for the third benchmark. Computational results with the three
140 benchmarks are discussed below.

141 In the first level parallelism, a spatial (global or limited-area) domain is divided
142 into M different sub-domains with one or two-dimensional domain decomposition,
143 running with M MPI processes in parallel. The “pseudo codes” for the first-level
144 parallelism, which is performed in the spatial domain, is shown in List 2. Here, we
145 first present the performance of PEEMD code in analyzing the ERSST data, which is
146 on a 2-dimensional (2D) spatial grid with 180 and 89 grid points in longitude and
147 latitude, respectively. The 2-D domain is decomposed into M subdomains by
148 dividing the longitude interval into I subintervals and the latitude interval into J
149 subintervals, where $M=IxJ$. Table 1 shows the performance of the code in processing
150 the ERSST data using the first parallelism with various values of (I, J). The second
151 row with $(I, J) = (1, 1)$ indicates the serial run that is used as the baseline for the
152 calculation of the speedup. The remaining rows show the speed-up using only the
153 first level MPI parallelization. Since data size is small for this case, only a small

154 number of MPI processes was used. As shown in the Table 1, the overall scalability
155 is good with a speedup of 10.1 using 16 cores, yielding a parallel efficiency of 0.63.
156

157 In each subdomain decomposed by the first level parallelism, the second level
158 parallelism is deployed to decompose the total number of ensemble trials into N
159 groups, running with N MPI processes in parallel. Then, the third level parallelism
160 applies the fine-grain OpenMP threads within each of the N processes that were
161 spawn by each of the M tasks. To examine the performance of OpenMP threading,
162 Table 2 shows the numerical performance of one Sandy Bridges node, which has 16
163 cores, using various number of OpenMP threads. While the number of the MPI
164 processes is fixed at four, which is used to limit the communication among the MPI
165 processors, the number of threads is increased from one, two, to four in order to
166 examine the performance of the OpenMP threading. The second column of Table 2
167 shows the speedup of 1.82 and 3.08 using 2 and 4 threads, respectively. In this
168 calculation, a traditional Box-Muller transformation is employed in the random
169 number generation routine. Having realized the disadvantages of Box-Muller
170 transformation, random number generation of the code has been greatly improved by
171 using Intel's Vector Statistical Library (VSL), which is part of the Intel Math Kernel
172 Library (MKL, <https://software.intel.com/en-us/articles/intel-math-kernel-library-documentation>). The third column of Table 2 shows the improved scalability, with a
173 speedup of 1.97 and 3.81 using 2 and 4 threads, respectively.
174

175 As our ultimate goal is to apply the PEEMD to analyzing global model
176 simulations at a resolution of 1/4 degree or higher, the performance of the three-level
177 parallelism in the PEEMD is first discussed using the idealized MRG wave solution
178 with a number of grid points comparable to that at the target model resolution. The
179 number of grid points is 1001x1001 for the MRG wave case, while the global
180 weather/climate model at a resolution of 0.25 degree (e.g., Shen 2012a) has

182 1001x721 grid points. The PEEMD with such as grid resolution required 400MB of
183 memory per variable. Table 3 presents computational results from the PEEMD with
184 the 3-level parallelism using up to 5000 Ivy Bridges cores. The last two columns of
185 the Table 3 show the values of (I, J) that are used to divide the 2D domain into IxJ
186 subdomains, giving a total number of subdomains $M=IxJ$. The third column “Ens” in
187 the Table displays the number of the MPI processes for the second level parallelism
188 that is applied in the ensemble space, while the 4th column “OMP” shows a number
189 of OpenMP threads within each of the MPI process in the second level parallelism.
190 We selected the case with 60 cores as the baseline, because of required memory in
191 this specific application. As shown in Table 3, the PEEMD is scalable up to 5,000
192 cores. In particular, we obtained a parallel speedup and performance of 52.8 and
193 63%, respectively, by increasing the number of cores from 60 to 5000. Figure 1
194 displays the log-log plot of the scaling as a function of the number of cores. (Please
195 add more discussions here.)

196

197 In summary, major features of the newly-developed PEEMD include: (1)
198 promising scalability that yields a parallel speedup and efficiency of 52.8 and 63% by
199 increasing the number of cores from 60 to 5000; (2) bit-by-bit consistency that
200 assures the binary identical results with different CPU layouts; and (3) sustainability
201 enabled by a well-designed interface, making it easy to plug new versions of the
202 EMD into the software package of the PEEMD. The parallel performance and
203 implementation of an early version of the PEEMD and its application were discussed
204 in Shen et al. (2012b) and Cheung et al. (2013). The scientific performance of the
205 PEEMD is discussed in the next section.

206

207 3 Application for the analysis of Earth science data

208

209 To verify the scientific performance of the PEEMD, we apply the PEEMD to
 210 analyzing the following three meteorological datasets: (1) a chaotic solution of
 211 Lorenz model (1963); (2) an idealized Kelvin cat's eye flow (e.g., Smith, 1986; Shen
 212 and Lin, 1998); and (3) Hurricane Sandy (2012) (Shen et al., 2013).

213

214 3.1 The solution of the Lorenz model

215

216 The first case is designed to understand the performance of the PEEMD in the
 217 analysis of solutions' envelop with the aim of understanding the corresponding
 218 growth rate and period at different stages of the nonlinear solution. To achieve our
 219 goal, we apply the tool to performing the envelope analysis of the solutions near a
 220 non-trivial critical point in the Lorenz model (Lorenz, 1963). The Lorenz model with
 221 three state variables, which is referred to as the 3DLM, is first linearized with respect
 222 to its non-trivial critical point solutions (e.g., page 135 of Lorenz, 1963; Eq. 21 of
 223 Shen, 2014), as follows:

$$\frac{dX'}{d\tau} = -\sigma X' + \sigma Y', \quad (3)$$

$$\frac{dY'}{d\tau} = (r - Z_c)X' - Y' - X_c Z' - FN(X'Z'), \quad (4)$$

$$\frac{dZ'}{d\tau} = Y_c X' + X_c Y' - b Z' + FN(X'Y'), \quad (5)$$

224 here, τ is the dimensionless time. We use $\tau=110$ and $b=8/3$ in this study. σ and r are
 225 Prandtl number and normalized Rayleigh number, which are set to be 10 and 25,
 226 respectively. The flag FN indicates if the system is fully nonlinear ($FN=1$) or "linear"
 227 ($FN=0$). Each variable with a prime (e.g., X') indicates perturbation with respect to
 228 its critical point (X_c), and the total field is written as $X = X_c + X'$. In the above
 229 equations, the choice of the critical point solutions as the basic state leads to the
 230 disappearance of the nonlinear terms that involve the interaction of two basic state
 231 variables (e.g., $X_c Y_c$). In other words, the explicit terms with the product of two

variables only contain the interaction between one basic state (e.g., X_c or Y_c) and one perturbation field (e.g., Y' or X') and the interaction between two perturbation fields (e.g., X' and Y'). The interaction between two perturbation fields ($X'Y'$ or $X'Z'$) is small when the initial perturbation (e.g., departure from its critical point) is small. By neglecting these terms (i.e., $FN=0$), Eqs. (3-5) become linear with respect to the critical point solution. The linear system should be able to depict the initial evolution of the solution in the full 3DLM when initial perturbations are small, which is shown below.

240

241 The (local) growth rate and period near the critical point of the 3DLM are
242 estimated using the solution's envelope (i.e., consecutive maximum). Numerical
243 solutions are obtained with the total integration time of $\tau=110$ and an integration time
244 step $\Delta\tau=0.001$. Initial conditions are $(X', Y', Z') = (0, 1, 0)$. Shown in Figure 2 are the
245 local growth rates and periods using the numerical solutions with $FN=1$ and $FN=0$,
246 which are indicated by blue and red open circles, respectively. The theoretical
247 eigenvalue of $\lambda = (7.92109 \times 10^{-3}, i 9.67213)$ obtained from the eigenvalue problem of
248 the 3DLM (e.g., Eq. C8 of Shen, 2014) gives the linear growth rate of 0.0079 and
249 period of 0.6496 ($=2\pi/9.67213$), which are shown in black lines. A comparison in
250 Figure 2 suggests that the envelope analysis with the PEEMD produces the linear
251 growth rate and period in agreement with the theoretical values and that a (small)
252 initial perturbation grows at a linear grow rate initially and then at a nonlinear growth
253 rate.

254 3.2 A Kelvin cat's eye flow

255

256 Based on analysis of observations, Dunkerton, Montgomery, and Wang (2009)
257 proposed the marsupial paradigm, suggesting that a Kelvin cat's eye circulation may
258 appear in association with the critical level (or latitude, CL) of an easterly wave and
259 lead to the formation of a "pouch", which in turn provides protective environment for

260 small-scales convections to grow to enhance mesoscale (vortex) circulation and lead
261 to TC genesis. The CL is defined as the level where the phase speed of a wave mode
262 is equal to the speed of the environment flow. The second case is designed to detect
263 the Kelvin cat's eye flow and its association with tropical cyclogenesis.

264
265 With the confidence in the performance of the PEEMD in analyzing the idealized
266 cases such as the mixed Rossby gravity (MRG) wave (e.g., Wu and Shen 2015, under
267 review), the second data set with grid points of (1001, 1001) in the (X, Y) directions
268 are created as follows. Horizontal and vertical wind perturbations (U' and W') are
269 generated using analytical solutions of a terrain-induced large-scale flow (e.g., Smith
270 1986; Eqs. 12-13 of Shen and Lin, 1999). The basic wind (U) represents a linear
271 back-sheared flow, $U=U_0(1-Z/H)$, where U_0 represents the surface wind, Z is the
272 height and H is height of wind reversal. These solutions with the basic wind display
273 the so-called Kevin cat's eye flow in Figure 2 of Smith (1986), showing the
274 streamline from the total horizontal wind ($U'+U$) and vertical wind (W'). Individual
275 components of horizontal winds (U' , and U) and the total horizontal wind ($U'+U$) are
276 shown in Figure 3a-c, respectively. The total wind (Figure 3c) is then decomposed
277 into nine IMFs using the PEEMD. Only the sixth IMF (IMF6), the ninth IMF (IMF9)
278 and their summation (IMF6+IMF9) are displayed in the middle panels of Figure 3.
279 These IMFs are selected because of high correlation coefficients between each of the
280 IMFs and wind perturbation (or basic wind). Note that the IMF9 is indeed the non-
281 oscillatory residual of the data. It is clear that the sixth IMF (IMF6 in Figure 3d) and
282 ninth IMF (IMF9 in Figure 3e) represent well the wind perturbation (Figure 3a) and
283 basic wind (Figure 3b), respectively. In addition, we obtain a very high correlation
284 coefficient of nearly 1 (0.999) between the U' (U) and the IMF9 (IMF6), as indicated
285 by the scatter plots in Figure 3g-h. Encouraged by these promising results from the
286 idealized case and real world cases (e.g., Wu and Shen, 2015), we applied the
287 PEEMD to analyzing the environmental flows associated with Hurricane Sandy,
288 which is discussed below.

289

290 3.3 Hurricane Sandy (2012)

291

292 For our third case, we apply the PEEMD to analyzing Hurricane Sandy (2012)
293 data. Numerical approaches and analysis results are given below. Storm Sandy first
294 appeared as a tropical storm in the southern Caribbean Sea on October 22, 2012,
295 moved northeastward, turned northwestward, and made landfall near Brigantine,
296 New Jersey in late October. Sandy devastated surrounding areas, caused an estimated
297 damage of \$65 billion, and became the second costliest tropical cyclone (TC) in the
298 U.S. history, surpassed only by Hurricane Katrina (2005). Previously, Shen et al.
299 (2013) demonstrated the model's capability to realistically predict Sandy's genesis
300 with a leadtime of up to six days and subsequent evolution for the next two-day
301 period of October 22-24. They suggested that upper-level tropical waves (e.g., MRG
302 wave) and a low-level westerly wind belt may contribute to Sandy's formation.
303 Below, we demonstrate the performance of the PEEMD in revealing the upper-level
304 tropical waves.

305

306 Figure 4 shows the time longitude diagram of the upper level meridional winds
307 from the ERA-Interim reanalysis data (Dee et al., 2011) and the corresponding IMFs.
308 Among the IMFs, the fourth IMF (IMF4 in Figure 4c) displays a more visible wave
309 mode than the original data (Figure 4a). IMF4 is then used to determine what kind of
310 waves may appear. In panel (c), black dashed lines, indicating the constant phase,
311 are used to determine the phase speeds, while a green line is used to estimate a
312 wavelength of 45 degrees. Theoretical dispersion relations (e.g., Matsuno, 1966;
313 Shen et al. 2012a and references therein) with this wavelength predict a period of
314 4.91 days for a mixed Rossby-gravity wave and 7.41 days for an equatorial Rossby
315 wave. The results suggest that the initial formation and movement of Hurricane
316 Sandy (as indicated by the cross signs in panel c) may be influenced by the mixed

317 Rossby-gravity wave and equatorial Rossby wave. Additional analysis with
318 traditional methods can be found in Shen et al. (2013).

319 **4 Summary**

320
321 In this study, we discussed the implementation of the parallelism in the PEEMD
322 and presented computational and scientific results with the PEEMD on the Pleiades
323 supercomputer. The three-level parallelism in the PEEMD leads to a parallel speedup
324 and efficiency of 52.8 and 63% by increasing the number of cores from 60 to 5,000.
325 The remarkable scientific performance of the PEEMD was illustrated with three
326 different cases to (i) perform an envelope analysis to calculate the growth rate and
327 period for the chaotic solution of the nonlinear Lorenz model (1963); (ii) extract a
328 specific wind component from an idealized Kelvin cat's eye flow constructed using
329 analytical solutions; and (iii) analyze multiscale tropical waves associated with
330 Hurricane Sandy (2012). Further improvement in the computational and scientific
331 performance is being made. In addition, the PEEMD will be used in a multi-year data
332 analysis to reveal the statistical relationship among different scale weather and
333 climate.

334

335 **Acknowledgments:**

336 We are grateful to the following organizations for their support: the College of
337 Science at San Diego State University and the NASA Advanced Information Systems
338 Technology Program. Resources supporting this work were provided by the NASA
339 High-End Computing Program and the NASA Advanced Supercomputing Division
340 at Ames Research Center.

341

342 **References:**

343

- 344 Cheung, S., B.-W. Shen, P. Mehrotra , J.-L. F. Li, 2013: Parallelization of the
345 Ensemble Empirical Model Decomposition (PEEMD) Method on Multi- and
346 Many-core Processors. AGU 2013 Fall Meeting, San Francisco, December 9-13.
- 347 Dee, D. P., et al., (2011), The ERA-Interim reanalysis: configuration and
348 performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597.
349 doi: 10.1002/qj.828
- 350 Dunkerton T. J., M. T. Montgomery, and Z. Wang, 2009: Tropical cyclogenesis in a
351 tropical wave critical layer: Easterly waves, *Atmos. Chem. Phys.*, 9, 5587-5646.
- 352 Flandrin, P.; Rilling, G.; Gonçalves, P., 2003: Empirical Mode Decomposition as a
353 Filterbank. *IEEE Signal Processing Letters* 11 (2): 112–114.
354 doi:10.1109/LSP.2003.821662.
- 355 Lorenz, E. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20, 130–141.
- 356 Matsuno, T. (1966), Quasi-geostrophic motions in the equatorial area, *J. Meteorol.*
357 *Soc. Jpn.*, 44, 25–43.
- 358 Huang, N.E., Shen, Z., Long, S., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung,
359 C.C., Liu, H.H., "The Empirical Mode Decomposition and Hilbert Spectrum for
360 nonlinear and non-stationary time series analysis", Proc. R. Soc. London A, 454,
361 (1998), 903-995.
- 362 Smith, R.-B, 1986: Further development of a theory of lee cyclogenesis. *J. Atmos.*
363 *Sci.*, 43, 1582–1602.
- 364 Smith TM, Reynolds RW, Peterson TC, Lawrimore J, 2008: Improvements to
365 NOAA's historical merged land–ocean surface temperature analysis (1880–2006).
366 *J. Climate.*, 21:2283–2296.
- 367 Shen, B.-W., 2014: Nonlinear Feedback in a Five-dimensional Lorenz Model. *J. of*
368 *Atmos. Sci.* 71, 1701–1723. doi: <http://dx.doi.org/10.1175/JAS-D-13-0223.1>
- 369 Shen, B.- W., and Y.-L. Lin, 1999: Effects of Critical Levels on Two-Dimensional
370 Backsheared Flow over an Isolated Mountain Ridge on an f-plane. *J. of Atmos.*
371 *Sci.*, 56, 3286-3302.

- 372 Shen, B.-W., W.-K. Tao, and Y.-L. Lin, and A. Laing, 2012a: Genesis of Twin
373 Tropical Cyclones as Revealed by a Global Mesoscale Model: The Role of Mixed
374 Rossby Gravity Waves. J. Geophys. Res. 117, D13114,
375 doi:10.1029/2012JD017450.
- 376 Shen, B.-W., Z. Wu, and S. Cheung, 2012b: Analysis of Tropical Cyclones and
377 Tropical Waves using the Parallel Ensemble Empirical Model Decomposition
378 (EEMD) Method. AGU 2012 Fall Meeting, San Francisco, December 03-07,
379 2012.
- 380 Shen, B.-W., M. DeMaria, J.-L. F. Li and S. Cheung, 2013: Genesis of Hurricane
381 Sandy (2012) simulated with a global mesoscale model, Geophys. Res. Lett., 40,
382 4944–4950, doi:10.1002/grl
- 383 Wang, G., X.-Y. Chen, F.-L. Qiao, Z. Wu, and N. E. Huang, 2010: On Intrinsic Mode
384 Function. Advances in Adaptive Data Analysis. 2, 277-293.
- 385 Wu, Y.-L., and B.-W. Shen, 2015: An Evaluation of the Parallel Ensemble Empirical
386 Mode Decomposition Method in Revealing the Role of Downscaling
387 Processes Associated with African Easterly Waves in Tropical Cyclone
388 Genesis Atmos. Oceanic Technol. 33, 1611-1628, DOI: 10.1175/JTECH-D-15-
389 0257.1
- 390 Wu, Z. and N. E. Huang, 2004: Statistical significance test of intrinsic mode
391 functions. In *Hilbert-Huang Transform: Introduction and Applications*, pp 125-
392 148, Ed. N. E. Huang and S. S. P. Shen, World Scientific, Singapore, 311pp.
- 393 Wu, Z. H., and N. E. Huang, 2009: Ensemble empirical mode decomposition: A
394 noise-assisted data analysis method. Advances in Adaptive Data Analysis, 1(1), 1-
395 41. [doi:10.1142/S1793536909000047].
- 396
- 397
- 398

399 **List 1:** Acronym

3DLM	Three-dimensional Lorenz model
CL	Critical Level
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
ERSST	Extended Reconstructed Sea Surface Temperature
IMFs	Intrinsic Mode Functions
MKL	(Intel) Math Kernel Library
MPI	Message Passing Interface
MRG	Mixed Rossby-gravity
PEEMD	Parallel Ensemble Empirical Mode Decomposition
TC	Tropical Cyclone
VSL	(Intel) Vector Statistical Library

400

401

402

403 **List 2:** Pseudo codes for the first level parallelism of the PEEMD

404

405

406

```
1 Integer, parameter :: imax=4000
2 Integer, parameter :: jmax=4000
3 Integer size, rank, kount
4 Integer i, ifirst, istart, iend, msize, mrank
! msize: the size of the group "m"
! mrank: the rank of a specific PE in the group "m"
!MPI initialization to set "size" and "rank"
5 allocate (indata(imax, jmax))
6 allocate (dataline(jmax))
7 allocate rslt(jmax, Nm+2)
8 allocate rsltld1(imax, jmax, Nm+1)
9 allocate buf(jmax, Nm+2, msize)

10 ifirst = (mrank)*msize + 1
11 kount=0

12 do i=ifist, imax, msize
13     dataline(:) = indate(i,:)
14     call mp_eemd(jmax, dataline, ..... , rslt)
15     call MPI_BARRIER(...)
16
17     call MPI_GATHER(rslt, jmax*(Nm+2), , buf, jmax*(Nm+2), ...)
18     if (rank .eq. 0) then
19         istart = 1+ knout * msize
20         iend = istart + msize - 1
21         do k=1, Nm+1
22             do j=1, jmax
23                 rsltld1(istart:iend, j, k) = buf(j, k+1, :)
24             enddo
25         enddo
26     endif
27     kount=kount+1
28     call MPI_BARRIER(...)
29 enddo
```

407 **Table 1:** The timing of the PEEMD with the first level parallelism in processing the
 408 ERSST data. The first two columns (I, J) display the values that are used to divide the
 409 spatial domain into IxJ subdomains. The third columns shows the timing in seconds
 410 and the fourth column displays a speedup that is calculated using the serial case with
 411 (I,J)=(1,1) as the baseline.

412

413

414

I	J	Timing (s)	Speedup
1	1	863.88	1.0
1	2	435.78	1.9824
2	2	257.90	3.3497
6	1	197.41	4.3760
4	2	142.43	6.0652
2	4	144.64	5.9727
4	3	111.63	7.7384
4	4	85.53	10.0997

415

416

417

418

419

420

421

422

423

424

425

426

427 Table 2: The parallel performance of the PEEMD associated with various number
428 of OpenMP threads. In this case, the number of MPI processes is fixed at four. The
429 first column shows the number of OpenMP threads. The second column displays the
430 time in seconds that were spent in the calculation using the original Box-Muller
431 Transformation for random number generation, while the third column shows the
time in seconds using the VSL library.

432

433

434

435

436

437

438

439

440

441

442

443

444

OMP	Original	VSL
1	146.52	135.92
2	80.32	68.94
4	47.55	35.66

445

446 Table 3: The performance of the PEEMD with the full three-level parallelism in
 447 processing the MRG wave case with up to 5,000 Ivy Bridges cores used. The first
 448 two columns (I, J) display the values used to decompose the 2-D spatial domain to M
 449 subdomains, yielding $M=IxJ$. Using the second level parallelism, each of the M
 450 processes spawns “Ens” (in the 3rd column) MPI processes in the ensemble space.
 451 The fourth column shows the number of OpenMP threads that are used in the third
 452 level parallelism. The total number of Ivy Bridge cores is shown in the 5th column.
 453 The timing, speedup, and parallel efficiency are shown in the last three columns.
 454 Note that the baseline case uses 60 cores.

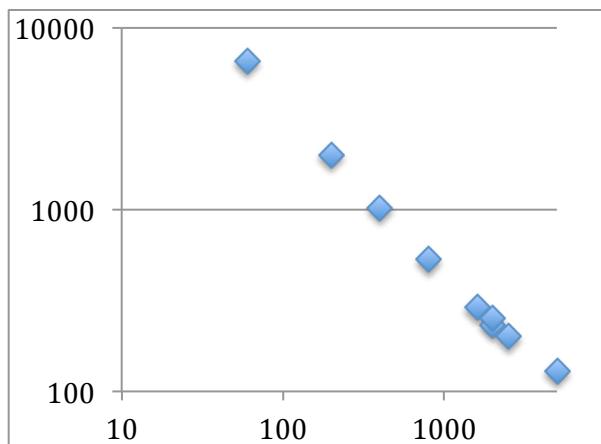
455

456

I	J	Ens	OMP	Total Cores	Time (s)	Speedup	Parallel Efficiency
5	6	2		60	6543.56	1	1
10	10	2		200	1983.25	3.3	0.99
10	10	4		400	1021.1	6.4	0.96
20	20	2		800	531.36	12.3	0.92
20	20	4		1600	289.42	22.6	0.85
25	40	2		2000	231.69	28.2	0.85
25	20	2	2	2000	251.21	26	0.78
25	25	4		2500	200.6	32.6	0.78
25	25	4	2	5000	129.68	50.4	0.60
50	50	2		5000	123.85	52.8	0.63

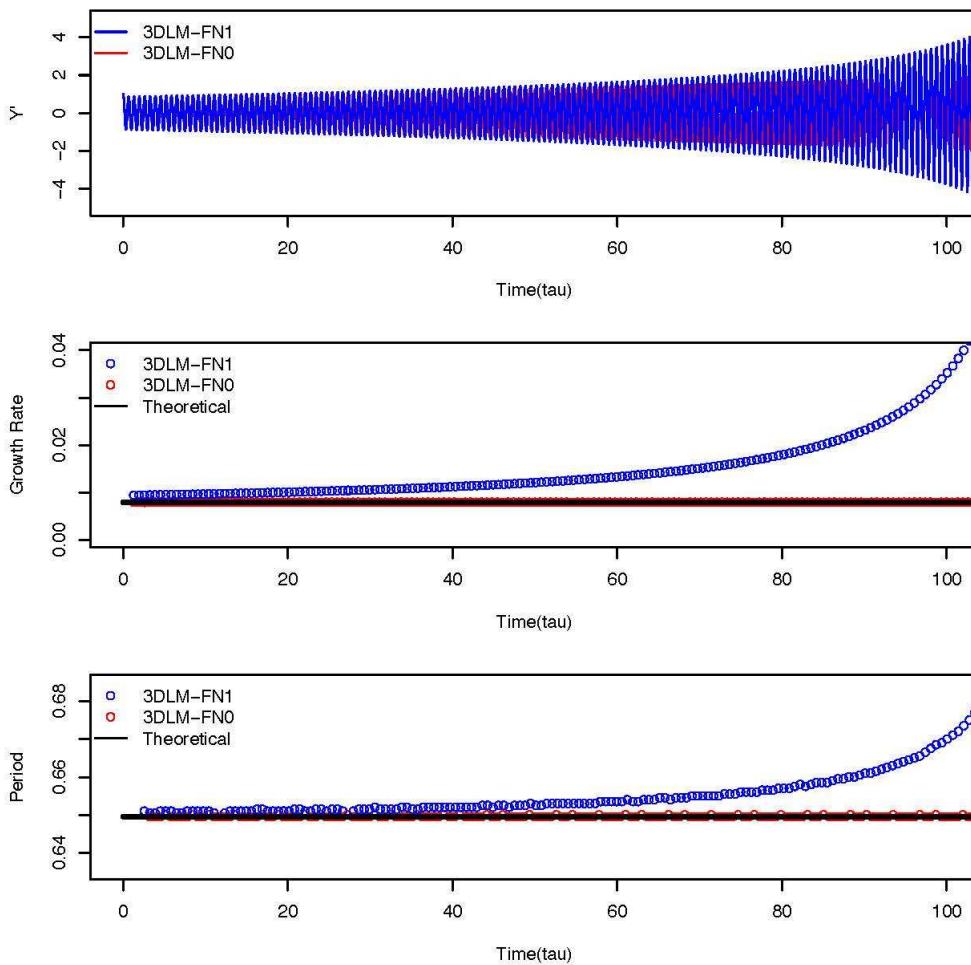
457

458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476



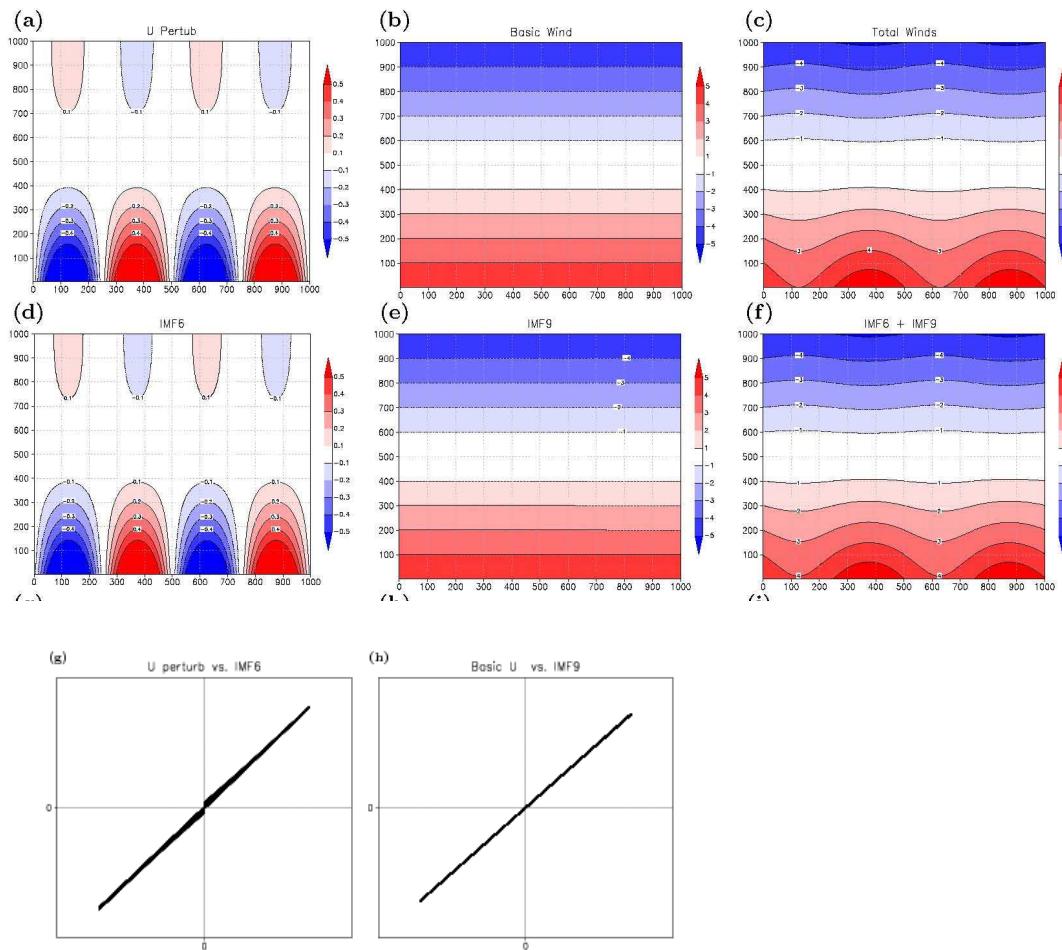
477 Figure 1: *Parallel performance of the PEEMD with three level parallelism in a log-*
478 *log plot. The vertical axis represents the time spent for the calculation and the*
479 *horizontal axis indicates a number of cores. The first case run with 60 cores takes*
480 *6,543.56 seconds while the last case run with 5,000 cores takes 123.85 seconds.*

481
482
483
484
485



486 Figure 2: *Local growth rate and period of the solutions' envelope in the 3DLM. The*
487 *blue and red open circles show the growth rates and periods from the numerical*
488 *solutions using the 3DLM described in Eqs. 3-5 with FN=1 and FN=0, respectively.*
489 *The black line shows the “theoretical values” using Eq. C8 of Shen (2014).*
490

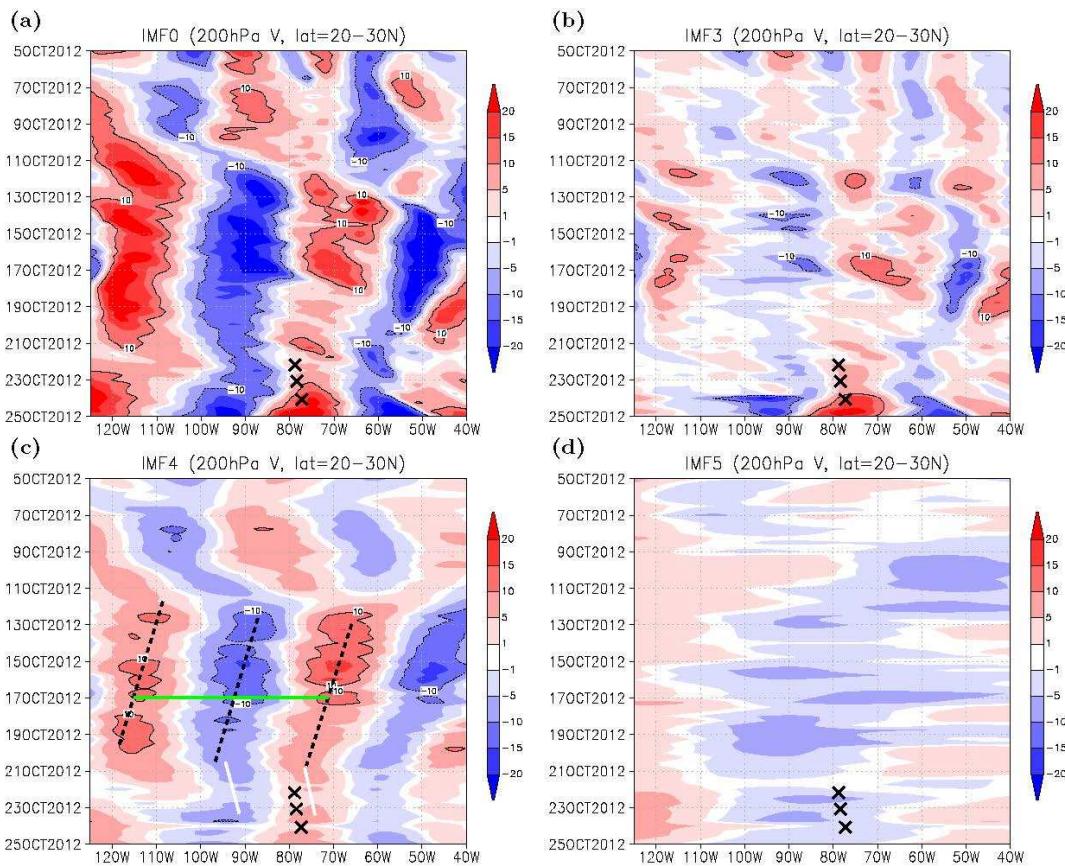
491



492 Figure 3: *Analysis of a Kelvin cat's eye flow using the PEEMD. Panels (a)-(c)*
493 *show the horizontal wind perturbation (U'), the basic shared flow (U) with a*
494 *wind reversal, and the total wind ($U'+U$), respectively. Panels (d)-(f) display*
495 *the decomposed components, including IMF6, IMF9, and their summation*
496 *(IMF6 + IMF9), respectively. Panels (g)-(h) show the scatter plot between the*
497 *IMF6 and wind perturbation U' ; and between the IMF9 and basic wind.*
498 *Analytical solutions in panel (a)-(c) are obtained from Smith (1986) (see also*
499 *Shen and Lin, 1999).*

500
501
502

503
504
505



506
507 *Figure 4: Decompositions of the upper-level (200-hPa) meridional winds from the*
508 *ERA-Interim reanalysis data. (a) Time-longitude diagram of meridian winds at 200-*
509 *hPa averaged over 20 and 30 degrees north latitude (reproduced from Shen et al.,*
510 *2013). (b, c, d) The third, fourth and fifth IMFs extracted from the 200-hPa*
511 *meridional winds, respectively. The locations of Sandy at 00Z Oct 22-24 are shown*
512 *in black multiplication signs.*

513
514
515
516