Reply to “Comment on “A self-consistent model of the interacting ring current ions and electromagnetic ion cyclotron waves, initial results: Waves and precipitation fluxes” and “Self-consistent model of the magnetospheric ring current and propagating electromagnetic ion cyclotron waves: Waves in multi-ion magnetosphere” by Khazanov et al. et al.”

G. V. Khazanov (1), K. V. Gamayunov (1), D. L. Gallagher (1), and J. U. Kozyra (2)

(1) NASA Marshall Space Flight Center, Huntsville

(2) Space Physics Laboratory, University of Michigan, Ann Arbor

1. Introduction

It is well-known that the effects of electromagnetic ion cyclotron (EMIC) waves on ring current (RC) ion and radiation belt (RB) electron dynamics strongly depend on such particle/wave characteristics as the phase-space distribution function, frequency, wave-normal angle, wave energy, and the form of wave spectral energy density. The consequence is that accurate modeling of EMIC waves and RC particles requires robust inclusion of the interdependent dynamics of wave growth/damping, wave propagation, and particles. Such a self-consistent model is being progressively developed by Khazanov et al. [2002, 2006, 2007]. This model is based on a system of coupled kinetic equations for the RC and EMIC wave power spectral density along with the ray tracing equations.

Thorne and Horne [2007] (hereafter referred to as TH2007) call the Khazanov et al. [2002, 2006] results into question in their Comment. The points in contention can be summarized as follows. TH2007 claim that: (1) “the important damping of waves by thermal heavy ions is completely ignored“, and Landau damping during resonant interaction with thermal electrons is not included in our model; (2) EMIC wave damping due to RC O⁺ is not included in our simulation; (3) non-linear processes limiting EMIC wave amplitude are not included in our model; (4) growth of the background fluctuations
to a physically significant amplitude “must occur during a single transit of the unstable region” with subsequent damping below bi-ion latitudes, and consequently “the bounce-averaged wave kinetic equation employed in the code contains a physically erroneous assumption”. Our reply will address each of these points as well as other criticisms mentioned in the Comment.

TH2007 are focused on two of our papers that are separated by four years. Significant progress in the self-consistent treatment of the RC-EMIC wave system has been achieved during those years. The paper by Khazanov et al. [2006] presents the latest version of our model, and in this Reply we refer mostly to this paper.

2. EMIC Wave Damping and Non-Linear Processes

EMIC wave damping due to thermal heavy ions and electrons has always been included in our studies (see Khazanov et al. [2002], Section 2; Khazanov et al. [2003], Section A3.2; Khazanov et al. [2006], Sections 3.1, 4.1, and 5.2; Khazanov et al. [2007]). In particular, an essential part of the study by Khazanov et al. [2007] is devoted to energy deposition to thermal plasmaspheric electrons due to Landau damping of EMIC waves. So the criticism listed in statement (1) in the Introduction of this Reply is simply incorrect.

During the main phase of magnetic storms, hot RC O+ may also contribute to strong damping of the He+-mode of EMIC waves [Thorne and Horne, 1997]. Although there is no doubt that this damping process is important, we have serious concern over the ability of the RC model used in the paper of Thorne and Horne [1997] to adequately represent the situation during the main and early recovery phase of a storm. Let us provide observational results that strongly support our statement. Braysy et al. [1998] reported observations of EMIC waves obtained by the Freja satellite, and provided remarkable results and conclusions. In particular, they observed oxygen band waves for about 7 hours during the later part of the main phase of the April 2-8, 1993 storm. Since the estimated drift time for RC O+ is only 2-4 hours, one would expect to find oxygen band
waves at different MLTs. However, all oxygen waves were found in the evening-midnight sector and, in particular, none were observed in the prenoon sector. This implies very asymmetric O⁺ RC during the main phase, and suggests that the RC oxygen ion loss rate is considerably faster than the drift speed. As emphasized by Braysy et al. [1998], these results are difficult to explain in terms of charge exchange and Coulomb scattering, and suggest that the production of EMIC waves contributes significantly to RC O⁺ decay during the main and early recovery phases. In other words, due to generation of the oxygen band EMIC waves, most RC O⁺ precipitates before reaching the dusk MLT sector.

The reported experimental facts clearly demonstrate that to adequately take into account He⁺-mode damping by RC O⁺ ions the O⁺-mode of EMIC waves should be also be included in global simulation. While O⁺-mode EMIC waves are not yet included in our model, we hope this will be completed in the near future. In any case, Table 1 in [Thorne and Horne, 1997] was obtained from a global simulation without oxygen band waves, and it is unlikely that the listed RC O⁺ parameters adequately represent the situation during the studied storm, especially at MLT=17 and L=4.25 for which all the calculations were presented. In addition, Thorne and Horne [1997] used bi-Maxwellian fits to the simulated RC distribution functions prior to calculating growth/damping rates, which will be shown next to incorrectly predict wave growth/damping and the resulting impact on the RC.

Next let us evaluate the possibility of excluding He⁺-mode damping by RC O⁺ for the May 2-7, 1998 storm that was studied in our published papers [Khazanov et al., 2006, 2007]. Using the RC kinetic model of Jordanova et al. [1998], Farrugia et al. [2003] found that RC O⁺ content did not exceed 30% during the main phase of this storm. Note that this estimate was obtained from a global simulation similar to that used by Thorne and Horne [1997] and did not include oxygen band waves. Therefore, as discussed above, Farrugia et al. [2003] overestimated the RC O⁺ content during the event. On the other hand, the calculations of Thorne and Horne [1997] clearly confirm that the above RC O⁺ percentage cannot significantly support wave amplification, and only slightly
influences the resulting growth of He\(^+\)-mode. It is for this reason we chose to initially exclude RC O\(^+\) in our particular simulation of May 2-7, 1998. Therefore the criticism summarized in statement (2) above is misplaced relative to the studies of Khazanov et al. [2002, 2006].

Let us now address the criticism that non-linear processes that limit EMIC wave amplitude are excluded from our model. The non-linear interaction of large amplitude EMIC waves (e.g., the modulational instability that results in generation of solitons and is described by the derivative nonlinear Schrodinger equation [e.g., Gamayunov and Khazanov, 1995, and references therein]) leads to phase correlation, and in such a system the wave-ion interaction is quite different in comparison with a quasi-linear approach. Another mechanism of non-linear EMIC wave saturation due to lower hybrid wave generation with subsequent Landau damping on thermal plasma was discussed by Gamayunov et al. [1992] and Khazanov et al. [1997, 2004]. In order to describe the latter saturation process, a full kinetic particle-in-cell (PIC) code should be used (see initial results by Singh and Khazanov, [2004, 2007]). Because the hybrid model suggested for use by TH2007 treats electrons as a massless fluid it is not an appropriate tool for modeling all possible EMIC wave saturation processes. During further development of our model, the various possibilities for strong non-linear wave-particle and/or wave-wave interaction will be taken into account, as needed, by using, for example, PIC simulation (see Singh and Khazanov, [2004, 2007] for more details).

At present, our model is based on quasi-linear equations, and the validity of the quasi-linear approach has been monitored from the first version of the model (see equation (8) and following text in Khazanov et al. [2003]). The quasi-linear validity criterion employed in our model is based on the test particle simulation of Kuramitsu and Hada [2000] who showed when quasi-linear diffusion is consistent with non-linear diffusion. Note that quasi-linear EMIC wave saturation takes place during most of the storm time, and the introduced criterion only restricts wave energy during the main and early recovery phases. Consequently, the criticism summarized in statement (3) above is incorrect.
3. EMIC Wave Propagation and Amplification
3.1. Theoretical Considerations

The TH2007 statement that growth of background fluctuations to physically significant amplitude must occur during a single transit of the unstable region is based on the calculations of Thorne and Horne [1997]. First, let us note that we are able to reproduce the Thorne and Horne [1997] results for path-integrated gain (not show here) using our code and their modeling parameters. In the paper of Thorne and Horne [1997], the RC H+ and O+ phase distributions were obtained from a Michigan RC-Atmosphere interaction Model (RAM) [Kozyra et al., 1997] simulation of the November 1993 magnetic storm, and then fitted by bi-Maxwellian distribution functions (see Table 1 in their paper). Unfortunately, it is not clear whether the Michigan RAM particle distributions in their paper were obtained with or without feedback from He+-mode of EMIC waves (in other words, whether the RAM simulation included wave-ion scattering or did not). Let us examine these two possibilities using results from the Michigan RAM and our model.

We assume the calculations presented below are rather general and should not depend on a particular storm. Below we refer to the May 2-7, 1998 magnetic storm (see Khazanov et al. [2006] for more details). Without EMIC wave damping by thermal plasma (which is normally included in our RC-EMIC wave model), we calculate the maximum equatorial growth rate for He+-mode EMIC waves. Figure 1 shows the ratio of the local equatorial growth rate calculated using a bi-Maxwellian fit to a simulated RC H+ distribution function, $\gamma_{bi-Max}$, to a corresponding growth rate obtained from the simulated distribution function without approximation, $\gamma_{num}$. Note that Figure 1 shows the ratios in locations where $\gamma_{bi-Max} > 10^{-2} \text{s}^{-1}$ only. This figure was generated at 80 hours after 0000 UT on 1 May 1998 during the early recovery phase. The left plate shows results from the
Michigan RAM without including wave feedback, the central plate corresponds to the Michigan RAM results with an empirical wave model included as described by TH2007 in the Section 2 of the Comment (see also [Kozyra et al., 1997]), and the right plate represents the results from our RC-EMIC wave model. The corresponding EMIC wave distributions from our model and the Michigan RAM can be found in [Khazanov et al., 2006, Figures 6 and 8]. Without wave feedback, the bi-Maxwellian fit most often overestimates the local equatorial growth rate by at least a factor of two. This overestimation increases dramatically (up to a factor of 30) if RC-EMIC wave scattering is included in the global simulation. The local RC H\(^+\) distribution function is not only affected by the local wave distribution but also depends on the prehistory of the storm. The wave feedback depends strongly on the EMIC wave model used, where there is an overestimation by a factor 9 for the Michigan RAM, and by a factor 30 for our model. Moreover, EMIC wave feedback can even cause \(\gamma_{num}\) to be negative while \(\gamma_{bi-Max} > 0\) (see gray color in Figure 1). Although the ratio \(\gamma_{bi-Max}/\gamma_{num}\) shown in Figure 1 is only less than zero for our model, this is also typical for the Michigan RAM simulation during the storm progress (not shown).

Figure 1 clearly demonstrates that use of a bi-Maxwellian fitted distribution for calculating growth rate routinely overestimates value predictions, especially during and/or after gamogenetically active periods when EMIC wave feedback is crucial for the fine structure of the RC distribution. This conclusion is not only true for hour 80 shown Figure 1 but for the entire May 2-7, 1998 storm simulation (not shown). The presented theoretical result is strongly supported by the observations of Anderson et al. [1996a]. These authors analyzed the proton cyclotron instability in the Earth’s outer magnetosphere, L>7, using Active Magnetosphere Particle Tracer Explorers/Charge Composition Explorer (AMPTE/CCE) magnetic field, ion, and plasma wave data. They found that magnetospheric hot proton distributions, from 1 to 50 keV, are not well characterized by a single bi-Maxwellian distribution. By fitting a sum of several bi-Maxwellians to the data, Anderson et al. [1996a] improved the analytical fit to the observations, reducing the residual between the fit and the data by factor of 4 to 30. The conclusion is that determination of \(T_{\perp}\) and \(T_{\parallel}\) by moment calculation is inadequate for
EMIC wave instability analysis [Anderson et al., 1996a]. As a result, the full RC distribution function on global magnetospheric spatial and temporal scales is required for a realistic assessment of wave excitation, and consequently, the overall wave effect on the RC population.

It is shown above that use of a bi-Maxwellian fit to the RC $H^+$ distribution can overestimate the local equatorial growth rate by an order of magnitude. The factor 10 was obtained using the Michigan RAM simulation, and re-evaluation using the full kinetic particle distribution is sufficient to decrease the wave gain that was otherwise obtained by Thorne and Horne [1997] to values well below what is needed to account for wave growth during a single transit of the unstable region.

The $He^+$-mode of EMIC waves is well guided along a magnetic field line and experiences “fast” quasi-periodic bouncing between surfaces of the $O^+-He^+$ bi-ion hybrid frequency in opposite hemispheres. [Doesn’t TH2007 call this into question using the Loto'aniu et al., 2005] paper? That would mean you can’t simply state it to be the case. It has to be substantiated first.] The wave normal angle of these waves oscillates about $\theta = \pi/2$ and progressively, but “slowly”, goes to 90 degrees. The ray path for the $He^+$-mode wave in the vicinity of the plasmapause, that is the most favorable region for EMIC growth, was illustrated in Figure A5 of Khazanov et al. [2006], and the timescales for “fast” and “slow” motions were found to be $\tau_{\text{fast}} \sim 10^2 \text{s}$ and $\tau_{\text{slow}} \sim 10^3 \text{s}$, respectively. Another timescale characterizing the wave evolution is a typical growth time and this was estimated to be $\tau_{\text{growth}} = 1/\gamma \sim 10^3 \text{s}$. Note that $\gamma$ includes both the energy source due to interaction with the hot RC and the energy sink due to absorption by thermal and hot plasmas and must be evaluated on a time scale of the wave bounce period [Khazanov et al., 2006]. The presented time scale hierarchy, along with the above theoretical and observational evidences, suggests that the bounce-averaged approximation employed in our RC-EMIC wave model is valid.

3.2. Observational Considerations
In order to achieve sufficient wave growth during a single transit of the unstable region, Thorne and Horne [1997] suggested a model in which the wave normal angle is confined to less than a 10 degrees cone centered on the magnetic field line over the entire ray path through the unstable region. This is crucial to obtain the wave gain reported by Thorne and Horne [1997]. The restriction suggests that EMIC wave ellipticity is close to -1 near the magnetic equator (left-hand polarized waves). This expectation contradicts observations [e. g., Fraser and Nguyen, 2001; Meredith et al., 2003; Loto’aniu et al., 2005]. Observations find that wave events in the vicinity of the magnetic equator are evenly distributed from left-hand polarized to near linear polarized and there is a clear tendency for the polarization to become more linear with increasing magnetic latitude. Because it is not applicable in this case, the observation of a significant number of linearly polarized events near the equator cannot be explained by polarization reversal from left-handed through linear to right-handed at the crossover frequency, as discussed for other events by Young et al. [1981] (quasi-field aligned waves can have a linear polarization if the Young’s mechanism works). Therefore the observed linear polarization suggests that waves will often be highly oblique inside the unstable region near the equator. Recently, using the more reliable wave step polarization technique, Anderson et al. [1996b] and Denton et al. [1996] analyzed data from the AMPTE/CCE spacecraft, and presented the first analysis of nearly linear polarized waves for which the polarization properties have been determined. They indeed found a significant number of wave intervals with the wave normal angle $\theta_{kB} > 70^0$. The above observations cannot be reconciled with the wave amplification model of Thorne and Horne [1997] ($\theta_{kB} < 10^0$ in their scenario) but have a natural explanation in the framework of our RC-EMIC wave model. The results presented by Khazanov et al. [2007] demonstrate that occurrences of the oblique and field-aligned wave normal angle distributions appear to be nearly equal near the magnetic equator with slight dominance of oblique events, consistent with observations.

Now we consider the TH2007 statement that the observational study by Loto’aniu et al. [2005] is consistent with the theoretical prediction of He$^+$-mode growth and damping by Thorne and Horne [1997], and successfully invalidates the concept of wave packet
bouncing between off-equatorial magnetic latitudes corresponding to the ion-ion hybrid frequency. *Loto'aniu et al.* [2005] used magnetic and electric field data from CRRES to obtain the Poynting vector for Pc 1 EMIC waves. They found bidirectional wave energy propagation, both away and toward the equator, for 26% of the events observed below $11^\circ$ [MLat], and unidirectional energy propagation away from the equator for all events outside $\pm 11^\circ$ of the equator. *Engebretson et al.* [2005] found a similar EMIC wave energy propagation dependence but with mixed direction within approximately $\pm 20^\circ$ [MLat], and consistently toward the ionosphere for higher magnetic latitudes. These observations lead *Engebretson et al.* [2007] to conclusion that "the mixed directions observed in the above studies near the equator is evidence of wave reflection at the off-equatorial magnetic latitudes corresponding to the ion-ion hybrid frequency. Waves that reflect would then set up a standing (bidirectional) pattern in the equatorial magnetosphere. Waves that tunnel through would tend to be absorbed in the ionosphere and not be able to return to equatorial latitudes." This conclusion by *Engebretson et al.* [2007] is in agreement with the physical picture underlying our RC-EMIC wave model and will next be shown to be consistent with the *Loto'aniu et al.* [2005] observations.

The statistical results presented by *Loto'aniu et al.* [2005] show that most of the unidirectional events outside $\pm 11^\circ$ of the equator are actually observed outside $\pm 18^\circ$ of the equator and a data gap between $-18^\circ$ and $-14^\circ$ is apparent (the data gap in the northern hemisphere is an orbital effect). *Loto'aniu et al.* [2005] estimate the bi-ion frequency location at $|\text{MLat}| = 15^\circ - 20^\circ$, which is consistent with the $10^\circ - 20^\circ$ [MLat] from *Rauch and Roux*, 1982; *Perraut et al.*, 1984. As a consequence, if there are heavy ions and waves are generated below the corresponding "bi-ion" latitude, they are able to tunnel through the reflection zone (or pass through this zone freely if waves are guided). Although there are no concurrent observations, let us consider spectrograms 3a and 3b from *Loto'aniu et al.* [2005] as typical. We can see that high latitude events have much less power than low latitude ones. This is consistent with tunneling from a low latitude source region to high latitudes through the bi-ion frequency. (Note that low frequency events shown in Figure 3a are more likely generated at high latitudes.) The implication is that waves are not strongly damped before/after reflection contrary to a remark by
TH2007. Inconsistency remains with identification of the transition latitude between bidirectional and unidirectional wave propagation in the two observational studies. However, the unavailability of wave observations at specific latitudes in [Loto’aniu et al., 2005] and/or differences in heavy ion content between the two studies may provide a resolution.

Observations presented by Loto’aniu et al. [2005] below 11° [MLat] show that 26% of the events support the concept of wave packets bouncing between the off-equatorial magnetic latitude corresponding to the ion-ion hybrid frequency. The events in Figure 3b of Loto’aniu et al. [2005] were observed at MLat = -10.5°, i.e. near the edge of the equatorial unstable region, and bidirectional wave energy propagation for packets b-h was observed. All these packets were mostly linear polarized and, as a result, waves were highly oblique. Note that, on average, simultaneous compressional Pc 5 wave amplitudes were less than 0.3 nT over the EMIC wave events, and it is unlikely that bidirectional pattern is due to Pc 5 modulation [Loto’aniu et al., 2005]. It is very difficult to generate highly oblique waves locally, and there is no active region below satellite location. So the equatorward wave packets are reflected below satellite at a latitude corresponding to O⁺-He⁺ bi-ion frequency. If this reflection point is located well below MLat = -10.5°, there is a conflict with the CRRES statistics because it did not observe the equatorially directed wave fluxes above 11° [MLat] but, as we pointed out above, this inconsistency may be due to unavailability of wave observations at specific latitudes in [Loto’aniu et al., 2005].

4. Responses to Other Comments

The TH2007 statement that “An implicit assumption for the applicability of equation (1) is that after reflection, wave energy is returned to the unstable region near the equator with propagation vectors aligned close to the ambient magnetic field direction to allow further amplification” is incorrect. Khazanov et al. [2006, equation (22)] explicitly included the ray tracing equations in RC-EMIC wave model. The growth/damping rate in
the right-hand side of the equation is a result of averaging of the local growth/damping rates along the ray phase trajectory \((r, \theta)\) over the wave bounce period, and the second term on the left-hand side of equation (22) takes into account the wave energy outflow from the region of small wave normal angles to \(\theta = 90^0\).

The TH2007 statement that “The approach used by [Khazanov et al., 2002, 2006], which allows bouncing waves to grow over 20 hours from the background level suggested by Akhiezer et al. [1975], is clearly not appropriate, since the wave growth is limited to a single transit of the equator” appears to result, in part, from a misunderstanding. The 20 hours stated in our work has nothing to do with the time period associated with wave growth. To start our simulation we construct the initial RC distribution using the statistically derived quiet time RC proton energy distribution of Sheldon and Hamilton [1993], and the initial pitch angle characteristics of Garcia and Spjeldvik [1985]. For EMIC waves, we use a thermal background noise level from Akhiezer et al. [1975]. The initial the RC and EMIC wave distributions are derived independently, and of course, have nothing to do with a particular state of the magnetosphere during a simulated event. Only the boundary conditions provided by the LANL satellites can be considered as data reflecting a particular geomagnetic situation (and, to a certain extent, the employed plasmasphere and electric field models driven by Kp). Therefore, we first seek an initial state for the RC and EMIC waves that is self-consistent and reflects the particular geomagnetic situation. In our case, this was done by running the model code for 20 hours before simulation of a particular event was possible. A similar preparation procedure should be done for any model and utilized initial data. Even if a perfect and complete RC and EMIC wave initial distribution was available from observations, they cannot strictly satisfy any set of governing equations (and so to any model) just because these equations do not include all real physics. Therefore again, 20 hours has nothing to do with the typical time for wave amplification and instead reflects the minimum time needed to adjust RC and waves to each other and to the real prehistory of a storm.

It is stated in TH2007, “Previous calculations of path-integrated wave gain during storm conditions [e.g., Thorne and Horne, 1997; Jordanova, 2005] are sufficient to drive waves
to the observed non-linear amplitudes during propagation through the unstable equatorial region". For this to be true, it is likely that the RC $H^+$ distribution must be subject to severe modification during less than half of the EMIC wave bounce period that is itself about the RC $H^+$ bounce period. This makes the RAM bounce-average RC formalism inadequate to obtain the RC distribution function.

5. Conclusions

The main points of this Reply can be summarized as follows:

(1) The EMIC wave damping by thermal heavy ions and electrons have always been included in all our studies.

(2) The RC $O^+$ can be neglected in the simulation of May 2-7, 1998 studied by Khazanov et al. [2006].

(3) Our model is based on quasi-linear equations and validity of this approach has been monitored in all versions of the model. Quasi-linear EMIC wave saturation takes place during the most of the storm time, and the controlling criterion restricts wave amplitude during the main and early recovery phases only.

(4) The insistence that wave growth takes place during a single transit of the unstable region is based on an approach (Thorne and Horne [1997]) that strongly overestimates growth rates as a consequence of approximating particle distributions with a bi-Maxwellian, it requires confinement of the wave normal angle of propagating waves to less than 10 degrees around the magnetic field, which contradicts observations, and violates the assumptions on which the bounce-averaged particle kinetic equation is based.

(5) The observation of EMIC waves by Loto'aniu et al. [2005] and Engebretson et al. [2005] at latitudes above the estimated reflection/tunneling points demonstrates that waves are not subject to severe damping before/after they get reflected/tunneled.

(6) Contrary to TH2007 statement that observational evidence contradicts our modeling results, the observations of Loto'aniu et al. [2005] and Engebretson et al. [2005, 2007] are consistent with our modeling and consistent with the only currently available explanation for the appearance of low latitude bi-directional EMIC wave propagation and higher latitude poleward directed unidirectional propagation. The explanation given by
Engebretson et al. [2007] is for a physical model of EMIC wave bouncing between the locations of the ion-ion hybrid frequency at conjugate latitudes with tunneling across the reflection zones and subsequent strong absorption in the ionosphere.

To conclude, we welcome this discussion because it draws focus to the details of what is needed to accurately model the RC-EMIC wave processes in Geospace. The issues raised by TH2007 represent important differences in long standing published research that need to be resolved before the community can coherently advance in this field. We maintain the validity of the RC-EMIC wave model published by Khazanov et al. [2006] through the discussion and evidence provided in this Reply.

References


Figure 1
May 1–7, 1998 Magnetic Storm, Hour 80
Ratio of Bi–Maxwellian Gamma to Exact Gamma

No waves

RAM

Our Model

\[ \frac{\gamma_{bi-max}}{\gamma_{exact}} \]