Dear Thomas, I think you wrote a nice paper. The related work section is sound.

A few comments and potential improvements regarding your analysis and methodology: - you propose the time fluctuation is due to network contention, as you primarily discuss & evaluate the impact of the compiler I think it would be important to really measure the time in the program for the computation only – this way an more accurate comparision of the compilers is possible. - when you mention the communication/computation ratio you could really measure it by using a profiler or trace environment like VampirTrace – this way you can really proof your assumption. - measuring the achieved FLOP/s and other hardware metrics might help further to assess the results. - For GCC you can/should try out the profile guided optimization (which is available in
your GCC), in many cases this helps to improve the performance significantly – I wonder what the performance will look like.

In the long term perspective I suggest to integrate an independent metric, which can be computed from the application directly – the simulation time/s is a good start, however it depends on the grid size. FLOP/s helps to understand how close WRF performance is to hardware performance – as a compiler might rewrite code to generate more floating point ops (if it assumes it is faster), thus this will allow to relate performance counters with theoretical values to understand compiler inefficiencies. Also, in absence of hardware counters the derived metric allows to assess runs with various problem sizes.

Regards, Julian

Interactive comment on Geosci. Model Dev. Discuss., 4, 547, 2011.