The authors master a wide range of techniques and build on a set of recently introduced methodologies to devise an interpolation scheme for forces derived from ab initio calculations. The manuscript appears to be technically correct. I have a number of comments on their general introduction and how they see their method in the context of other work, and a number of specific queries about the results.

General comments:

1) The assumption of locality at the bottom of page 2 underlies *all* interatomic potentials, including the general kernel based ones that the authors cite right above. This is acknowledged later in the middle of page 4. The selection of which ab-initio data is relevant is the role of the kernel that measures similarity of local environments.

2) The authors make much of the distinction between a local interpolation and a global interpolation, but this is just a technicality, because a global kernal-based interpolator can be converted to a local interpolator by employing cutoff on the similarity kernel. Using such a cutoff could end up being important in practice (or a significant drawback, introducing discontinuities into the dynamics), but is definitely not a fundamental aspect.

3) In the discussion the authors talk about "the tedious task of constructing a globally applicable empirical potential", but this is precisely what their work goes towards, automatically, and it's a jolly good thing too. Other groups are exploring this way of generating databases, although many have opted to add new entries in batches, rather than one by one. A closed-loop system that is truly fully automatic is a difficult beast to manage, and it is telling that after all the desires are spelled out, the authors stop short of carrying out their own programme, and they generate a database once and investigate its properties, just like all other groups.

4) In their development of the OGTO, when they relate it to SOAP, it is worth mentioning that in SOAP the (square of the) overlap is integrated over all rotations, whereas here the best match rotation is found. This choice is strange, because it leads to a discontinuity in the potential, because the best rotation can change discontinuously as one of the environments changes by a small amount. This is an insurmountable problem of the RMS-D type descriptors, but with overlap-based ones it can be circumvented by integration (rather than finding the best match), as in SOAP.

5) Eq 15 costs O(|BQ|^2) to solve. A global interpolation would cost O(database size). Which is larger in practice ? What are the typical numbers of configurations in BQ ?

6) In testing the method, the neighbourhood is defined in terms of "closest 4" or "closest 16" neighbours ? how does this generalise away from a crystal? what happenned in the 2500K simulations as the liquid flowed (Si is liquid at these temperatures, not an amorphous solid), is there a discontinuity introduced?

7) What is the computational cost of an OGTO-force? How does it compare to SOAP
(whose cost is on the order of 10-100 ms/atom for databases that are already useful)

Comments on results:

8) According to Fig 2b, even the 16-nearest neighbour model has expected error > 1 eV/A, resulting in quite a lousy model! How do the algorithms behave with more reasonable choices of 30-50 nearest neighbours that could produce force errors of 0.1 eV/A ?

9) Fig 9: The authors *must* show absolute force errors. just how good is this model? There are isolated points with no lines, why is that? Force is a vector, so what is being shown here, perhaps force error magnitude or force component error? I would prefer to see force component error.

10) How does the approach fare if one tries to include just two types of quite different different environments *simultaneously*, e.g. bulk and surfaces. The literature is full of papers describing complex methodologies that are tested on just bulk materials and showing "success", and that shows that bulk is easy.

11) Dynamics: how about computing an actual thermodynamic observable, e.g. angle distribution function as a function of temperature (near 2500K) or the phonon spectrum at 0 K (or low T). The radial distribution function is *not* enough, it is not a particularly sensitive mesure of model accuracy.

12) The title is certainly not justified: the degree of accuracy of the dynamics (i.e. trajectories and forces) was not assessed.

So overall, the authors set out on an ambitious programme, combining existing methodologies in a new way, and this is commendable. I don't think they have achieved anything particularly notable yet, and certainly do not show any results in the present version of the manuscript that would entice anyone to follow in their path, but by addressing my points above perhaps they could show that their approach improves or at least is on par with the state of the art.