Supporting information

Automatic generation of input files with optimised k-point meshes for Quantum Espresso self-consistent field single point total energy calculations

E. Patyukova*, J. Yin**, S. Basak, S. Pinilla Sanchez, A. Elena, G. Teobaldi

1 Goldilocks dataset

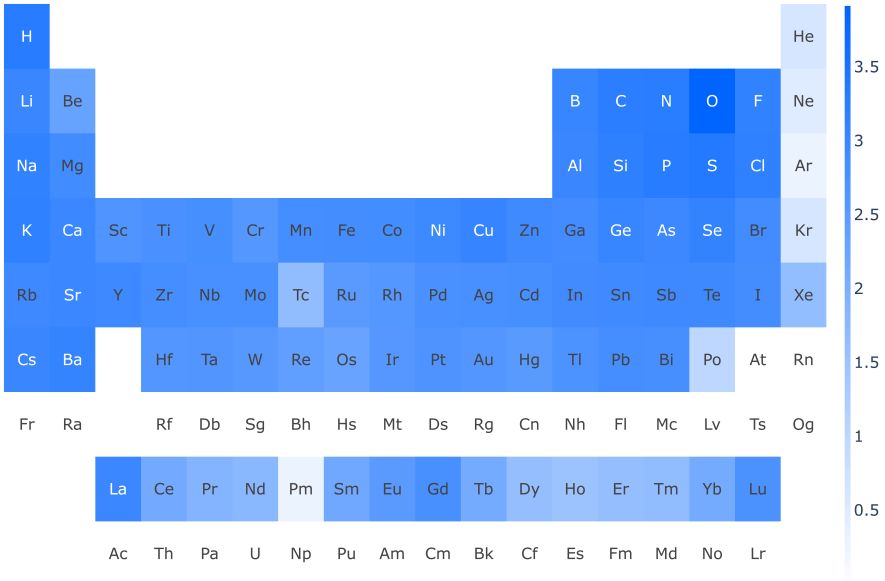

Figure S1: Representation of elements in the used dataset. Colour represents the log10 from the number of compounds containing this element.

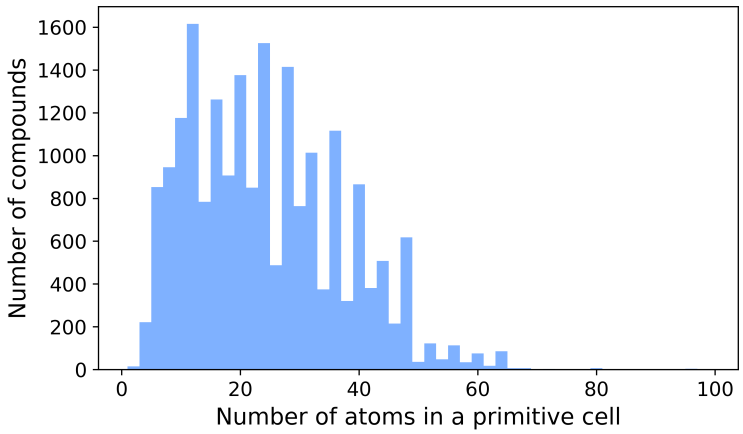

Figure S2: Distribution of the number of atoms in a primitive cell in the dataset

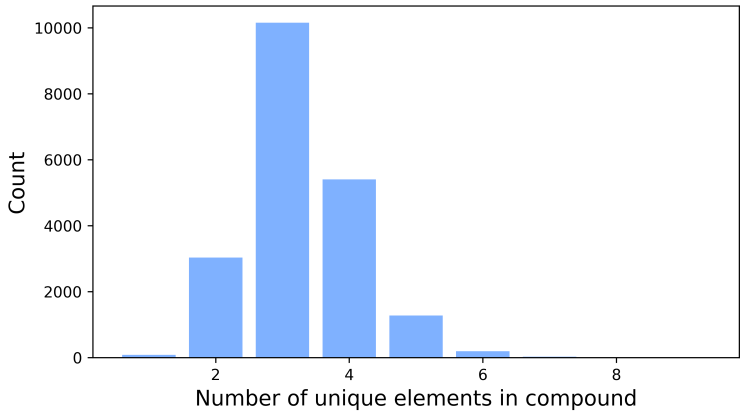

Figure S3: Distribution of compounds with respect to the number of elements in chemical formula

2 The front page of QE input generation application

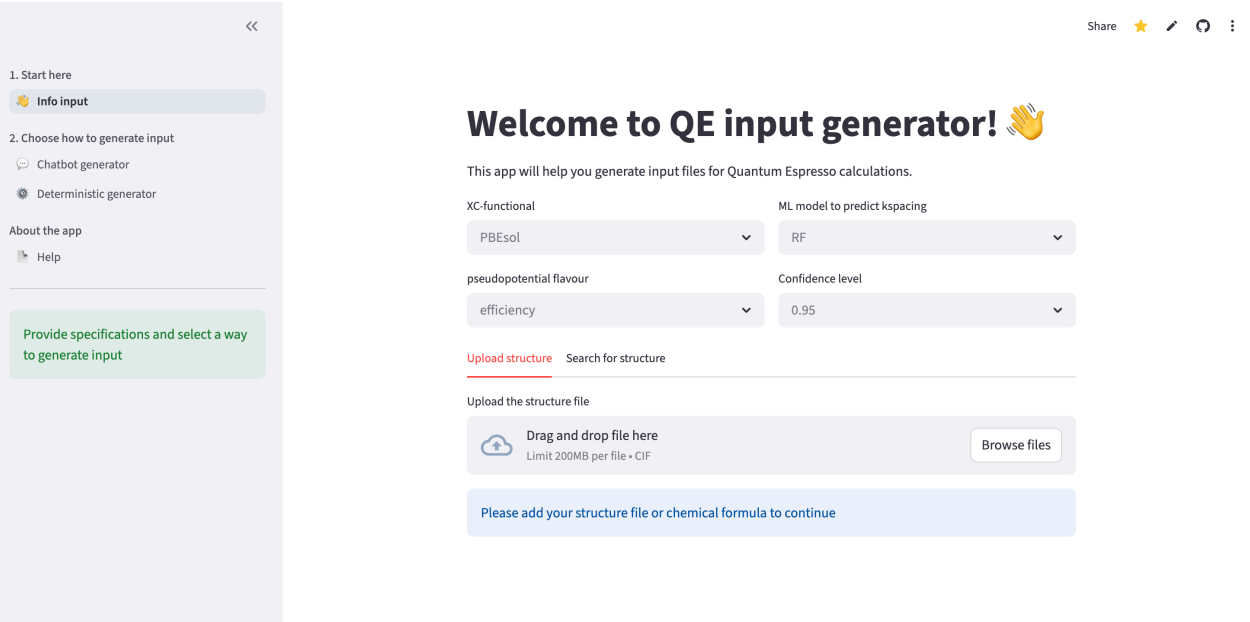

Figure S4: The front page of the application for generating input files for single-point SCF Quantum Espresso calculations

3 K-distance, K-mesh generation and convergence definition

Table 1: K-distance intervals and corresponding k-meshes for a representative material (MC3D source_db_id: S1929283).

k_dist_interval	k_mesh
[1.000, 0.714)	(1, 1, 1)
[0.714, 0.357)	(2, 2, 2)
[0.357, 0.238)	(3, 3, 3)
[0.238, 0.178)	(4, 4, 4)
[0.178, 0.142)	(5, 5, 5)
[0.142, 0.119)	(6, 6, 6)
[0.119, 0.102)	(7, 7, 7)
[0.102, 0.089)	(8, 8, 8)
[0.089, 0.079)	(9, 9, 9)
[0.079, 0.071)	(10, 10, 10)
[0.071, 0.064)	(11, 11, 11)
[0.064, 0.060)	(12, 12, 12)

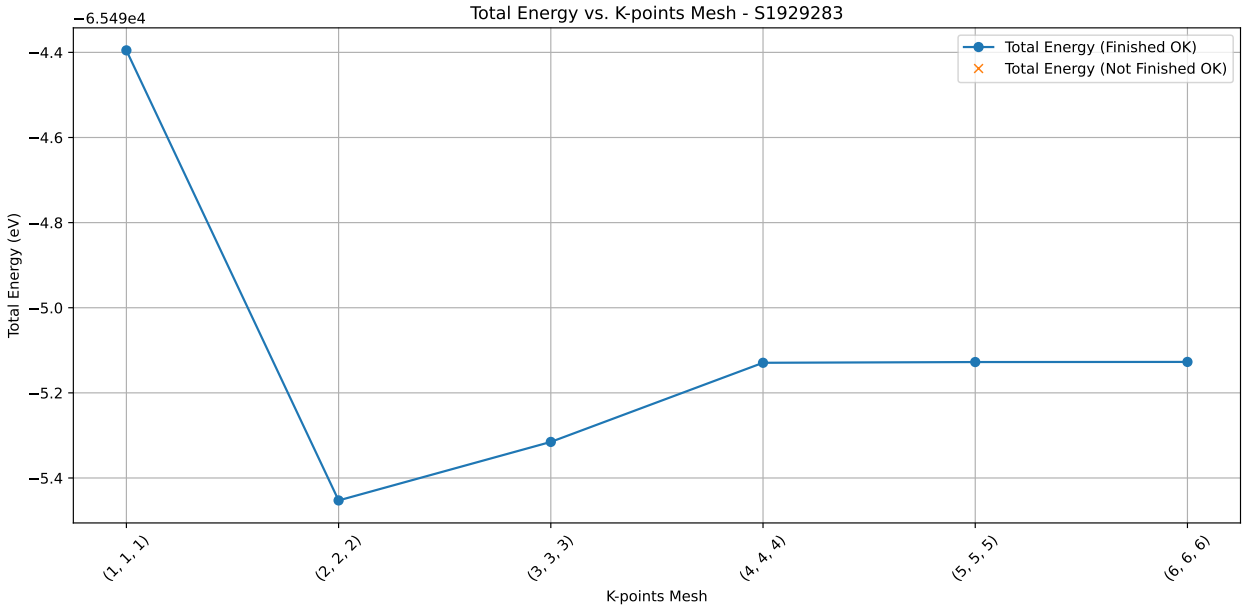

Figure S5: Total energy versus k-point mesh for material S1929283, showing convergence towards a stable value with increasing k-point density.

4 Comparison between the present Quantum ESPRESSO workflow and the JARVIS dataset

To assess the extent to which the convergence trends identified in the present work may transfer across computational workflows, we compared our converged k-distance labels with corresponding data from the JARVIS database for structures that could be uniquely matched between the two datasets.

A total of 6668 structures were uniquely matched. For each matched structure, we compared the JARVIS k-distance with the lower bound of the ultra-k-distance interval obtained in our present Quantum ESPRESSO workflow. The comparison is shown in Figure S6. The dashed line indicates the reference relation $y = x$.

Overall, the correlation between the two quantities is weak. Only 1.62% have a JARVIS k-distance lying within the corresponding ultra-k-distance interval defined in the present workflow, while 96.2% the JARVIS workflow yields a denser k-mesh than our converged result. This indicates that, although the underlying physical factors governing k-point convergence are expected to be general, the quantitative convergence labels remain strongly dependent on the specific computational workflow used to generate them.

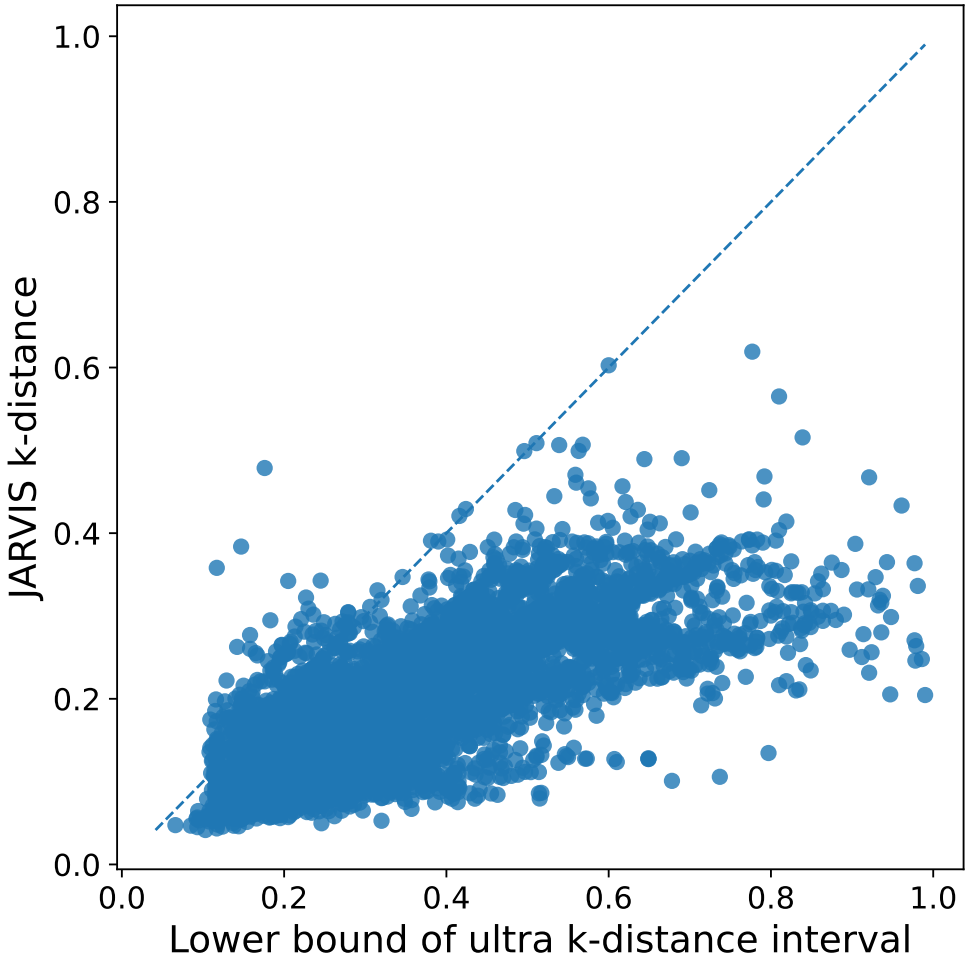

Figure S6: JARVIS k-distance plotted against the lower bound of the ultra-k-distance interval from the present Quantum ESPRESSO workflow for uniquely matched structures. The dashed line represents $y = x$.

5 Example of training the model to predict 1/k-distance

During the model development stage, we also tested alternative target variables, including 1/k-distance, which is more directly related to the effective density of the k-point grid. However, in our experiments the models trained to predict k-distance directly consistently achieved slightly better performance, as measured by the coefficient of determination (R^2) and MAE, compared to models trained on the transformed variable. For example, figure S7 shows RF+CSLM fitted on 1/k-distance.

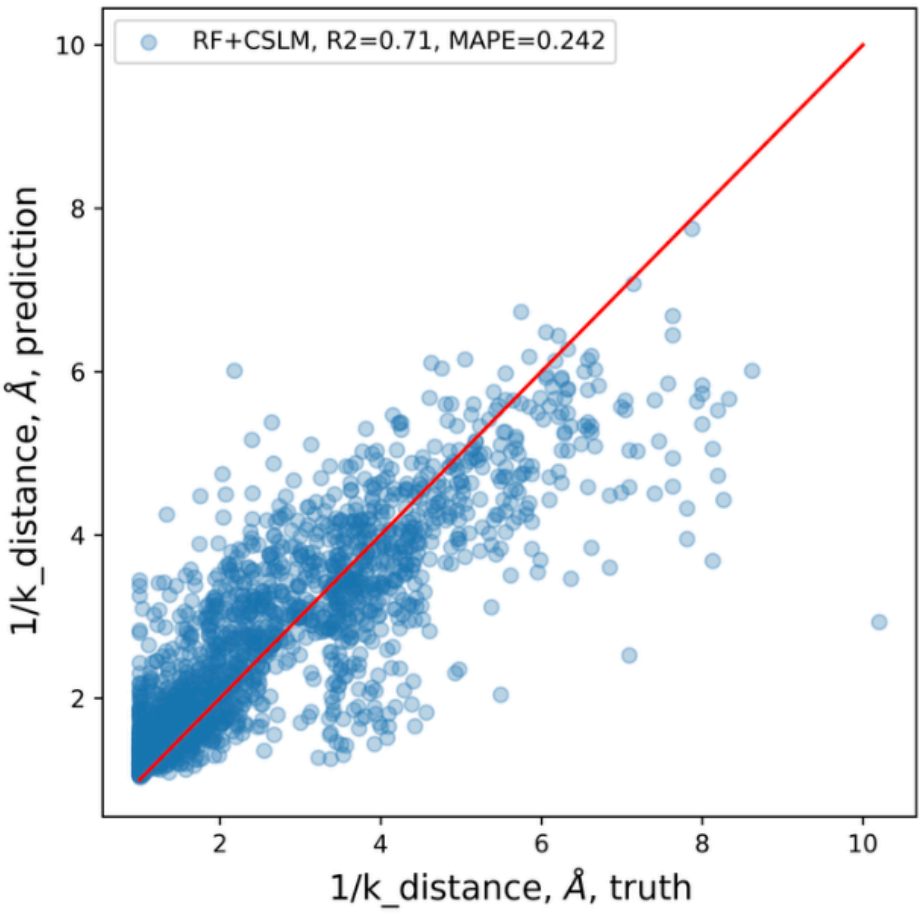

Figure S7: Predicted vs. actual 1/k-distance for RF+CSLM model trained on 1/k-distance.

6 Distribution of converged k-distance for metals and insulators

As discussed in the main text, the overall distribution of converged k-distance in our dataset is bimodal, which we attribute to the different convergence behaviour of metallic and insulating systems. To support this interpretation with explicit metallicity labels, we made use of the subset of structures that could be uniquely matched between our dataset and the JARVIS database (see Section 4 of this SI). For these matched structures, we adopted the metal/insulator classification reported in JARVIS, where a material is labelled as a metal if its electronic band gap is zero and as an insulator otherwise. The converged k-distances from our Quantum ESPRESSO workflow were then grouped according to this classification, and their distributions are shown in Figure S8.

The two groups exhibit clearly separated maxima: metals peak at a relatively dense k-distance of approximately 0.2 Å^{-1}, while insulators peak at a sparser k-distance of approximately 0.4–0.5 Å^{-1}. There is some overlap between the two distributions, which is expected since materials with small but finite band gaps can behave similarly to metals in terms of Brillouin-zone sampling requirements. This figure confirms that the bimodal shape of the overall k-distance distribution is largely driven by the underlying electronic character of the materials and supports the inclusion of metallicity-related features in the machine-learning models.

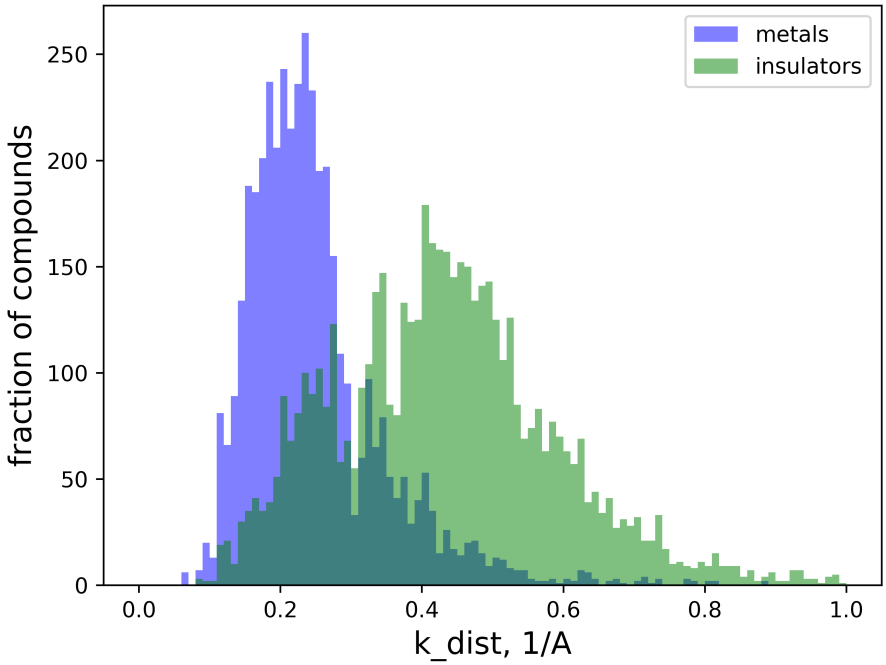

Figure S8: Distribution of converged k-distance for metallic (purple) and insulating (green) compounds, obtained from the subset of structures uniquely matched to the JARVIS database. The metal/insulator labels are taken from JARVIS, while the converged k-distances are those computed in the present Quantum ESPRESSO workflow.