**Cross-validation for spatiotemporal generalisability**

Since China lacked systematic ground-level measurements in earlier years before 2013, and the observation sites deployed in urban and rural environments were disproportional. We therefore decided to train the model at global scale with sufficient supervision by observations, and conducted strengthened rigorous cross-validation tests on the spatiotemporal extrapolation reliability to verify the generalisability of the deep learning downscaling algorithm. Besides the cross-validation and external validation tests by random split, we extended region-aggregated cross-validation tests on spatial extrapolation capability (cvs1: training on USA, testing on Europe; cvs2: training on Europe, testing on USA; cvs3: training on USA and Europe, testing on Asia; and cvs4: training on locations outside China, testing on China), and staged cross-validation tests on global-scale temporal generalisation (cvt1: training on 1990–2013, testing on 2014–2019; cvt2: training on 1990–2007 and 2014–2019, testing on 2008–2013; cvt3: training on 1990–2001 and 2008–2019, testing on 2002–2007; cvt4: training on 1990–1995 and 2002–2019, testing on 1996–2001; cvt5: training on 1996–2019, testing on 1990–1995) for the second-stage urban-rural differentiated downscaling.