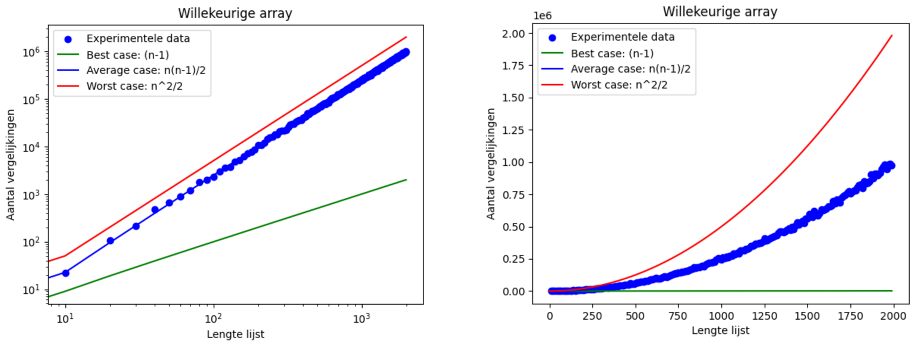

Bij het plotten van de data kan men kiezen voor een loglog plot (waarbij beiden assen op logaritmishe schaal staan) en een (normale) lineare plot. De loglog plot heeft het voordeel dat het de graad van de complexiteit duidelijk weergeeft in de richtingscoefficient van de grafiek. De lineare plot heeft dan weer als voordeel dat het intuitiever is voor een mens om in te schatten hoe de tijdscomplexiteit evolueert als functie van de invoer.

We kozen ervoor om arrays te onderzoeken met een lengte tot en met 2000. Enerzijds wil men zo groot mogelijke arrays onderzoeken aangezien dit meetfouten relatief verkleint. Anderzijds wil men de uitvoeringstijd van het programma haalbaar houden. 2000 leek ons daarbij een goed compromis.

Wanneer wel het programma laten lopen met integers duurde dit 6.9413 seconden, met floats was dit 6.2890 seconden. Best tegenintuitief dat de getallen met meer data sneller worden vergeleken. We schrijven het verschil in uitvoeringstijd toe aan een meetfout (misschien was er bij floating point een programma op de achtergrond aan het lopen).

Over het algemeen kunnen we besluiten dat de experimentele data enorm dicht bij het theoretische model aansluit. Er zijn kleine afwijkingen doordat een `willekeurige array` iets dichter bij een best-case of een worst-case kan ligge.

Wanneer we in bovenstaande grafiek de lagere orde termen zouden verwaarlozen, dan zouden we amper een verschil zien tussen beiden functies. Enkel in de loglog plot zou dit voor kleine lengtes van arrays een verschil geven. We kunnen besluiten dat we de lagere orde termen wel mogen verwaarlozen maar het moet zeker niet.

In de bovenstaande grafiek hebben we niet meerdere metingen per lengte van de array gedaan. Als we dit toch doen, zullen we zien dat de afwijkingingen op de average case zullen uitmiddelen.

Hoe kan je in de grafiek weergeven wat de spreiding is tussen het best-case en worst-case gedrag van InsertionSort? We stellen voor om niet meerdere arrays per lengte van array te gebruiken, zodat op de grafiek een grotere afwijking zichtbaar is. Daarnaast zou volgende formule ook een verband kunnen weergeven: $a = \frac{afwijking(best)}{afwijking(worst)}$ met $afwijking(f)$ de afstand tussen de datapunten (bijvoobeeld door middel van chi kwadraat). Wanneer a groter is dan 1, wijkt de data verder af van de best case en bijgevolg dichter bij de worst case. Omgekeerd geldt hetzelfde.

Het plotten van de theoretische boven- en ondergrenzen kan helpen om de lezer inzicht te geven in hoe dicht de average case zich bevindt ten opzichte van de best- en worst case. Echter, het is zeer belangrijk in het achter hoofd te houden dat de kans dat deze scenarios voorkomen bijzonder klein is. Immers, de kans is $\frac{1}{n!}$ wat voor grote n zo goed als nul is.