

L'Analyse de l'Aptitude d'un Procédé à l'Aide de MINITAB (II)

Keith M. Bower, M.S.

Résumé

Cet article développe le sujet traité dans le numéro précédent d'EXTRAOrdinary Sense. Les professionnels de la qualité constatent souvent que la supposition du modelage d'un procédé à l'aide de la distribution normale n'est pas valable. Cet article étudie deux approches pour aborder ce type de scénario, et propose un critère de prise de décision permettant de choisir entre les deux techniques pour une série spécifique de données.

Suppositions

Comme indiqué dans l'article précédent, deux suppositions critiques doivent être prises en compte lors de l'exécution d'analyses de l'aptitude d'un procédé, à savoir:

1. Le procédé est sous contrôle statistique.
2. La distribution du procédé étudié est Normale.

Comme le précise Walker¹, Abraham De Moivre a été le premier à découvrir la distribution normale en 1733. Il a depuis été constaté que cette distribution théorique pouvait bien modeler de nombreuses situations concrètes, y compris les distributions de hauteurs et de poids. Ceci est très important dans la théorie statistique, par exemple au niveau du Théorème de Limite Centrale (TLC). Bien que ce théorème nous aide dans la méthodologie des cartes de contrôle, il n'intervient pas dans les études de l'analyse de l'aptitude, quelles que soient les tailles d'échantillons obtenues.

Exemple

Prenons une distribution qui n'est pas normalement distribuée. N'oubliez pas que dans de très nombreuses situations, la supposition de Normalité n'est pas valable, par exemple lorsqu'il existe une frontière naturelle, disons à zéro, pour les données relatives au temps. Comme l'indiquent Somerville et Montgomery², les résultats obtenus à partir des analyses de l'aptitude basées sur la supposition de Normalité dans le cas effectif de distributions non normales peuvent conduire à de très graves erreurs.

Prenons les données de température à la figure 1. Dans cette simulation, nous supposerons une seule spécification, la limite inférieure, et cette LIS = 1° Celsius. Bien que le procédé soit sous contrôle statistique, la Droite de Henry met en doute la supposition de Normalité. L'histogramme au-dessus de cette Droite indique que la distribution est positivement asymétrique. Comme le montre le test d'Anderson-Darling (A-D) à la figure 2, nous pouvons rejeter l'hypothèse nulle selon laquelle les données suivent une distribution normale, même au niveau $\alpha = 0,01$.

¹ Walker, H.M., (1929). *Studies in the History of Statistical Method*, Williams and Wilkins, Baltimore, MD.

² Somerville, S.E., Montgomery, D.C., (1996). "Process Capability Indices and Non-Normal Distributions", *Quality Engineering*, Vol. 9.

Figure 1

Process Capability Sixpack pour Température

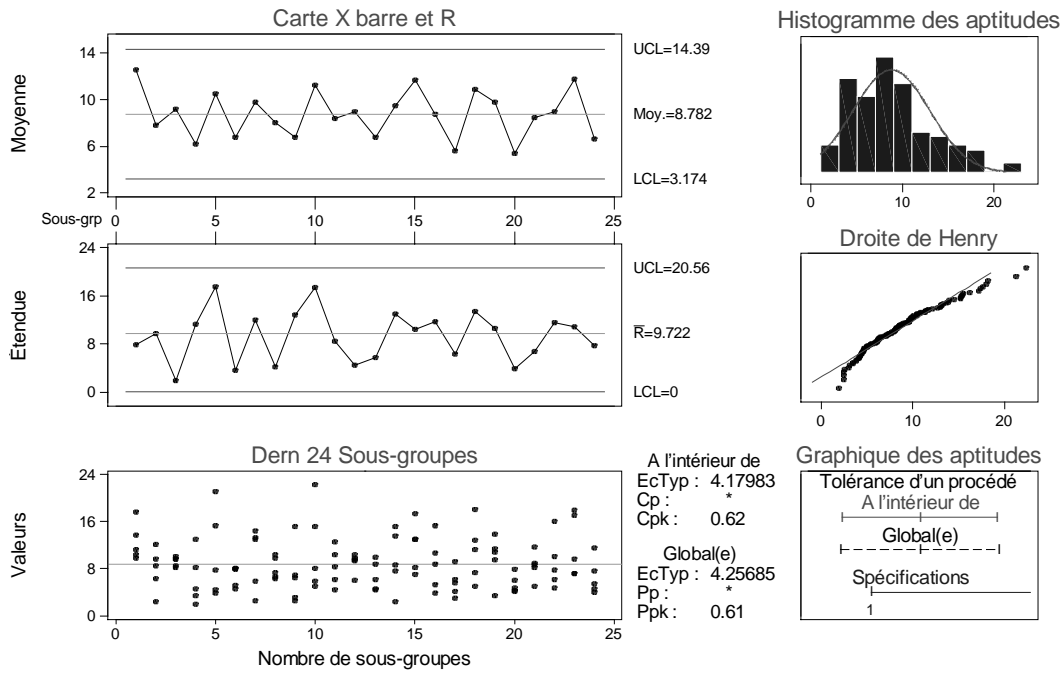
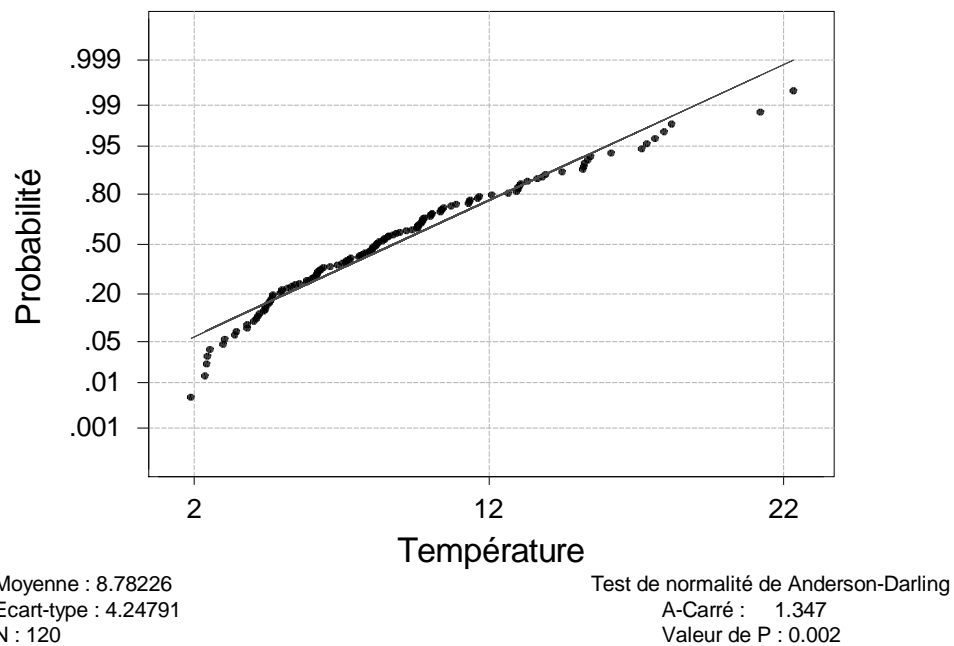


Figure 2

Graphique de la courbe normale ou droite de Henry



Pour reprendre l'article précédent, il est conseillé d'utiliser une Droite de Henry parallèlement à la valeur p associée au test d'A-D étant donné que ces valeurs p dépendent en grande partie de la taille de l'échantillon. Pour les échantillons importants,

vous rejetterez très probablement la supposition de Normalité, même si la distribution sous-jacente peut en fait être bien modélisée par la distribution normale.

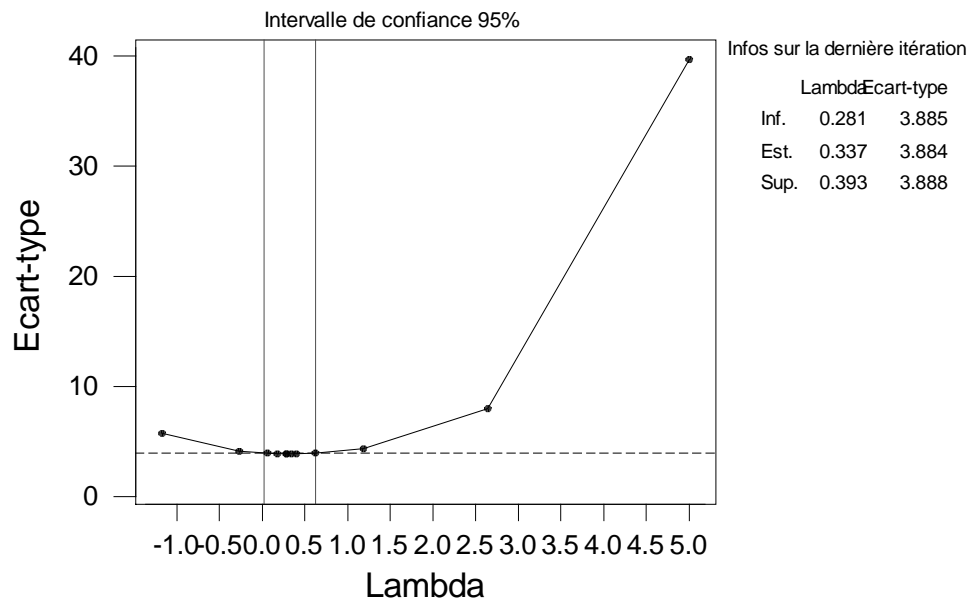
Parmi les options offertes pour exécuter une analyse de l'aptitude valable figurent:

1. La transformation des données, afin de pouvoir supposer à raison qu'elles suivent une distribution normale.
2. La supposition d'une autre distribution, telle que celle de Weibull.

Nous commencerons par étudier les opérations que nous pouvons effectuer pour transformer les données, notamment à l'aide de l'algorithme de Box-Cox³. L'algorithme même nous aide à déterminer le mode de transformation désiré des données (en élevant chaque observation à une puissance particulière), pour obtenir des valeurs qui, nous l'espérons, pourront être dûment modélisées par la distribution normale. Par exemple, l'utilisation de $\lambda = 2$ implique la mise au carré de chaque observation, tandis qu'une valeur de $\lambda = -1$ implique prendre l'inverse de chaque point de données. Nous déduisons d'après la figure 3 que la transformation optimum pour cette série de données consisterait à élever chaque observation à la puissance 0,337. Il convient toutefois de noter qu'il peut y avoir une différence pratique minimale entre l'utilisation d'une puissance de 0,337 et de 0,5, bien qu'une élévation à la puissance 0,5 (à savoir prendre les racines carrées) puisse être plus facile à comprendre et interpréter.

Figure 3

Diagramme de Box-Cox pour Température



Veillez noter dans ce cas que la valeur de 0,5 s'inscrit effectivement dans l'intervalle de confiance de 95% pour la valeur réelle de λ . Ceci suggère que la prise des racines carrées de la série de données d'origine puisse servir à obtenir une distribution quasi-normale. Si la valeur de 1 s'inscrivait dans cet intervalle, une transformation ne pourrait être justifiée.

³ Box, G.E.P.; Cox, D.R. (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society*, B, Vol. 26.

Après avoir étudié les mêmes données suite à leur transformation en prenant la racine carrée de chaque point, on constate que la transformation peut être justifiée. Veuillez noter que le test d'A-D affiche une valeur p supérieure à 0,10 (dans ce cas, la valeur p = 0,645 comme indiqué à la figure 4) et que la Droite de Henry ne présente aucun écart majeur de linéarité.

Nous pouvons par conséquent raisonnablement conclure que (i) le procédé est sous contrôle statistique et que (ii) les données transformées peuvent être supposées suivre approximativement une distribution normale. Les suppositions nécessaires semblent avoir été satisfaites et nous pouvons étudier l'aptitude de ce procédé, comme illustré à la figure 5.

Figure 4

Graphique de la courbe normale ou droite de Henry

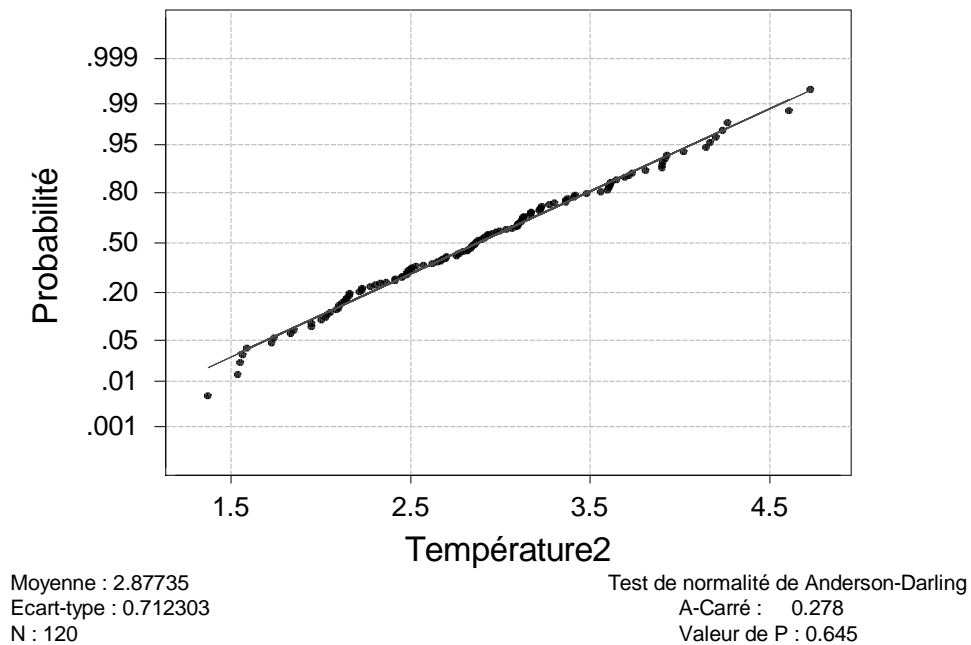
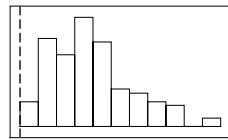


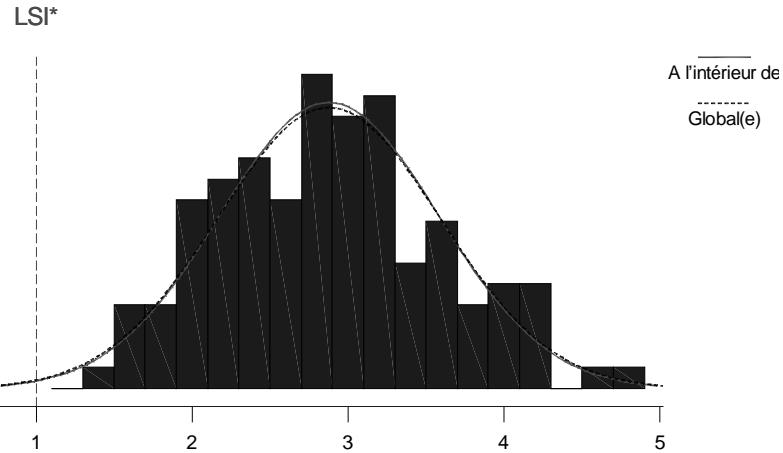
Figure 5



Analyse de l'aptitude du procédé pour Température

Transformation de Box-Cox avec Lambda = 0.5

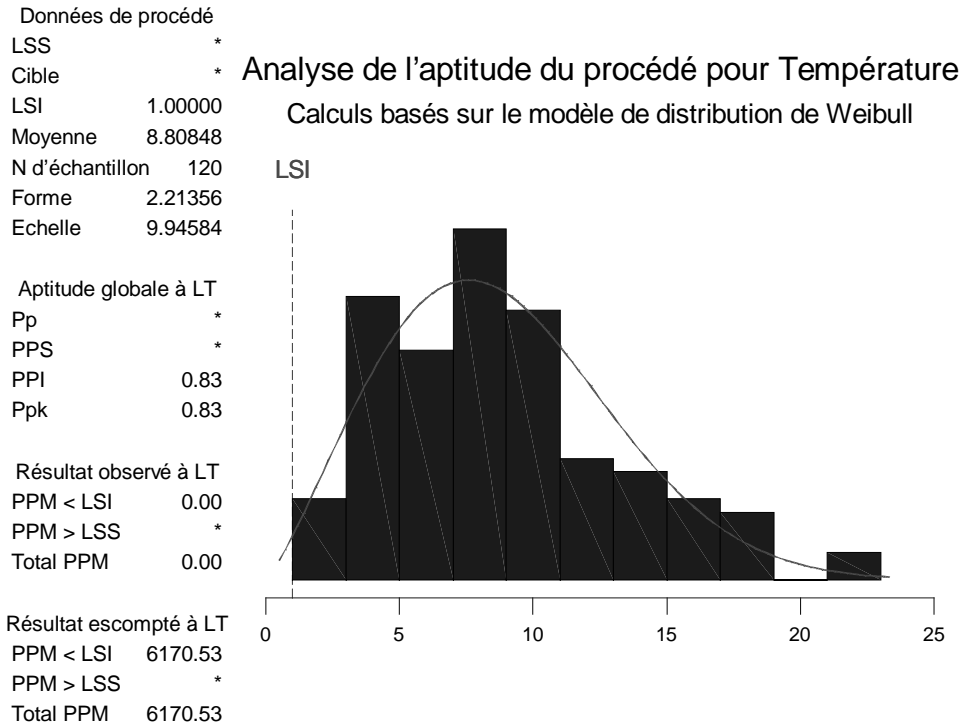
Données du procédé	
LSS	*
LSS*	*
Cible	*
Cible*	*
LSI	1.00000
LSI*	1.00000
Moyenne	8.78226
Moyenne*	2.87735
Echantillon N	120
EcT (à l'intér de)	4.17983
EcT* (à l'intér de)	0.70069
EcTyp (global)	4.25685
EcTyp* (global)	0.71380
Aptitude potentielle (à l'intér de)	
Z.référence	2.68
Z.LSS	*
Z.LSI	2.68
Cpk	0.89
Cpm	*
Aptitude globale	
Z.référence	2.63
Z.LSS	*
Z.LSI	2.63
Ppk	0.88



	Performances observées	Performances "à l'intér de" prévues	Performances "globales" prévues
PPM < LSI	0.00	PPM < LSI* 3688.95	PPM < LSI* 4268.39
PPM > LSS	*	PPM > LSS* *	PPM > LSS* *
PPM Total	0.00	PPM Total 3688.95	PPM Total 4268.39

Les statisticiens utilisent souvent la distribution de Weibull à des fins de modelage. L'analyse de l'aptitude à la figure 6 illustre l'analyse correspondante en supposant une distribution de Weibull pour la série de données d'origine. La question qui s'impose est donc: quelle est la meilleure technique pour notre analyse? Il est conseillé d'étudier la situation au cas par cas.

Figure 6



L'une des méthodologies possibles pour établir des comparaisons entre les deux procédures peut être d'étudier les graphiques de probabilité obtenus à l'aide des données transformées (en supposant une distribution normale) avec un graphique de probabilité de Weibull pour les données brutes. On peut également chercher à comparer les statistiques du test d'A-D en utilisant les ajustements de données normaux et de Weibull. La statistique d'Anderson-Darling indique l'éloignement des points relevés par rapport à la ligne ajustée sur un graphique de probabilité. La statistique même est un écart quadratique pondéré entre les points relevés et la ligne ajustée, ces poids étant plus importants aux extrémités de la distribution. Une statistique d'Anderson-Darling plus petite indique par conséquent que la distribution est mieux ajustée aux données. Il convient toutefois de ne pas se fier exclusivement aux statistiques d'A-D, car bien qu'une valeur puisse être inférieure, cela n'implique pas que l'ajustement du modèle est adéquat.

Comme illustré aux figures 7 et 8, et compte tenu des études précédentes, avec cette série de données, il peut être préférable d'utiliser les données transformées, et de supposer la distribution normale car la statistique d'A-D est inférieure ($0,379 < 0,548$), et nous avons constaté de manière explicite que nous ne pouvions pas rejeter l'hypothèse nulle au niveau 0,10 à la figure 4.

Figure 7

Normale Diagramme de probabilité pour Température2

Estimations MaxV - 95% IC

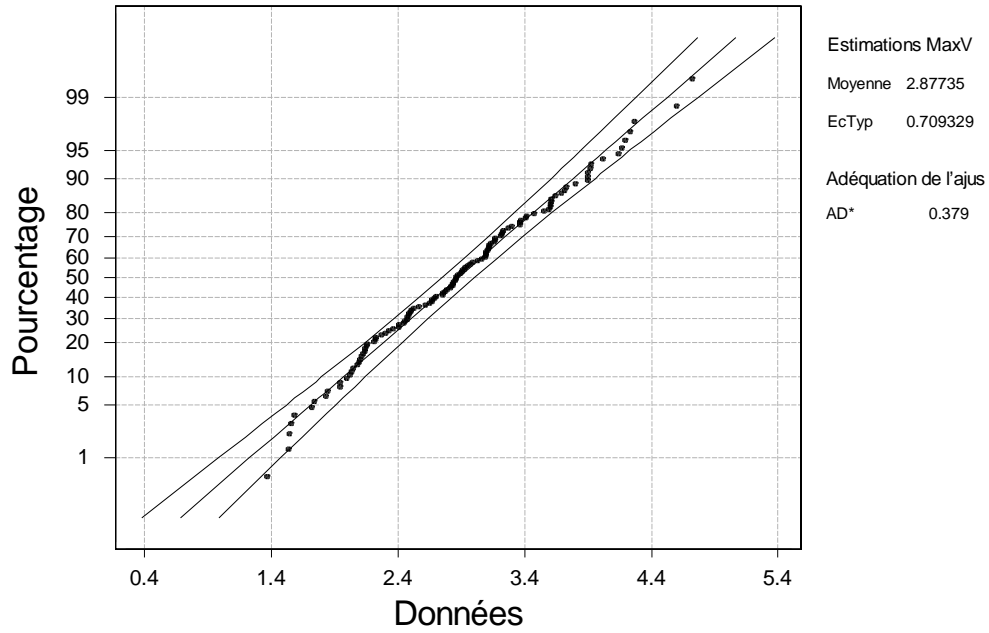
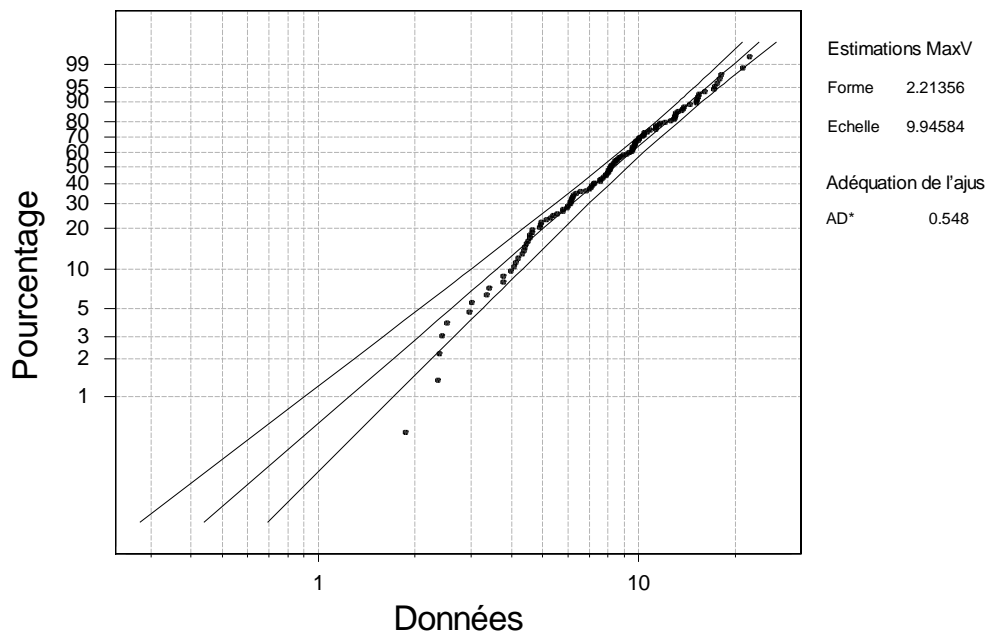


Figure 8

Weibull Diagramme de probabilité pour Température

Estimations MaxV - 95% IC



Conclusion

Des suppositions importantes sont associées au calcul des indices d'aptitude, qu'il s'agisse de C_p , C_{pk} , "Sigma", etc., et doivent être vérifiées avant de pouvoir utiliser ces statistiques en toute confiance. L'auteur espère que la prise en compte et la justification de ces aspects critiques, comme indiqué dans ces articles, pourront être utiles aux professionnels de la qualité en vue de l'exécution d'analyses de l'aptitude plus solides.

Keith M. Bower est titulaire d'une Maîtrise de Sciences, Gestion de la Qualité et Productivité, de l'Université de l'Iowa, et occupe les fonctions de Spécialiste de la Formation Technique chez Minitab, Inc. Les techniques d'amélioration continue de la qualité, notamment la réalisation de cartes de contrôle, ainsi que les méthodologies statistiques commerciales et de santé, constituent ses principaux centres d'intérêt.