Contributions to early-chatter detection and period-N bifurcations identification based on cumulative diagnosis approach

Yanqing Zhao

Ecole Doctorale IAEM (Informatique, Automatique, Electronique-Electrotechnique, Mathématiques et Sciences de l'Architecture)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention: «Automatique, Traitement du signal et des images, Génie informatique»

Contributions à la détection précoce de chatter et à l'identification des bifurcations de période-*N* basée sur une approche de diagnostic cumulatif

Par

Yanqing Zhao

Le 14 Décembre 2020

Membres du jury:

Rapporteurs:

M. Pierre DEHOMBREUX, Professeur, Université de Mons, Belgique.
M. Kamal MEDJAHER, Professeur, Université Fédérale Toulouse Midi-Pyrénées, France.

Examinateurs:

M. Ning HE, Professeur, Nanjing University of Aeronautics and Astronautics, Chine.
Mme Choubeila MAAOUI, Professeur, LCOMS, Université de Lorraine, France.
Mme Zeina AL MASRY, Maître de Conférences, ENSMM, FEMTO-ST, France.

M. Kondo H. ADJALLAH, Professeur, LCOMS, Université de Lorraine, France, Directeur.
M. Alexandre SAVA, Maître de Conférences HDR, LCOMS, Université de Lorraine, France, Co-directeur.

Laboratoire de Conception, Optimisation et Modélisation des Systèmes, LCOMS EA 7306
Université de Lorraine, Metz 57000, FRANCE

Ecole Doctorale IAEM (Informatique, Automatique, Electronique-Electrotechnique, Mathématiques et Sciences de l'Architecture)

Doctoral Thesis

presented and defended publicly for obtaining the grade of
DOCTOR DEGREE OF THE UNIVERSITE DE LORRAINE
in «Automatics, Signal and Image Processing, Computer Engineering»

Contributions to early-chatter detection and period-*N* bifurcations identification based on cumulative diagnosis approach

by

Yanqing Zhao

on 14 December 2020

Members of the jury:

Reviewers:

M. Pierre DEHOMBREUX, Professor, Université de Mons, Belgium.
M. Kamal MEDJAHER, Professor, Université Fédérale Toulouse Midi-Pyrénées, France.

Examiners:

M. Ning HE, Professor, Nanjing University of Aeronautics and Astronautics, China.
Mme Choubeila MAAOUI, Professor, LCOMS, Université de Lorraine, France.
Mme Zeina AL MASRY, Associate Professor, ENSMM, FEMTO-ST, France.

M. Kondo H. ADJALLAH, Professor, LCOMS, Université de Lorraine, France, Supervisor.
M. Alexandre SAVA, Associate Professor, LCOMS, Université de Lorraine, France, Co-supervisor.

Laboratoire de Conception, Optimisation et Modélisation des Systèmes, LCOMS EA 7306
Université de Lorraine, Metz 57000, FRANCE

Acknowledgments

I want to express my genuine gratitude to my supervisor, Professor Kondo Hloindo ADJALLAH, for his patience, guidance, and support during this journey. I appreciate his profound knowledge, scientific research approach, and unique perspective. Without his guidance and support, I will not finish this thesis.

I also want to express my sincere gratitude to my co-supervisor, Professor Alexandre SAVA, for his encouragement and help. His remarkable comments and suggestions help me grow up.

Also, I would like to thank my friends in Metz for their help and company. I have unforgettable memories with them. They are the wealth of my life.

Besides, special gratitude goes to my parents, wife, son, brother, and sister for their understanding, support, and encouragement. Undoubtedly, I cannot get a Ph.D. without their support and encouragement.

Finally, I thank the Huaiyin Institute of Technology for its support.

Résumé

Le diagnostic cumulatif des systèmes dynamiques nécessite la détection, l'identification et la caractérisation des dégradations naissantes. Son application à l'usinage à grande vitesse, par exemple, pourrait s'appuyer sur l'analyse des phénomènes de bifurcations de période-N pour détecter et identifier les chatters (broutages) naissants et améliorer la qualité des produits et des processus de fraisage. Jusqu'à présent, de nombreuses méthodes efficaces ont été proposées pour détecter les broutages naissants et identifier les bifurcations de période-N. Cependant, ces méthodes peinent à mettre en œuvre ces tâches de manière fiable et précise. Le but de la présente thèse est de développer et mettre en œuvre des méthodes de détection de broutages naissants et d'identification de bifurcations de période-N dans une approche de diagnostic cumulatif temps réel. Afin de détecter les défauts de broutages naissants (early-chatter), nous avons proposé trois méthodes de détection et une méthode d'identification pour le diagnostic cumulatif. La première méthode peut être utilisée pour détecter à distance les broutages naissants. La deuxième méthode détecte rapidement les broutages naissants dans des conditions spécifiques de fonctionnement et de mesure. Mais dans la pratique, les conditions de fonctionnement et de mesure sont complexes et variables. Pour s'adapter aux différentes conditions de fonctionnement et de mesure, nous avons proposé une troisième méthode et cette dernière détecte de manière fiable les broutages naissants. On note également que dans les processus de fraisage, les broutages peuvent naître avec une bifurcation de type période-N ou de type Hopf. La qualité d'usinage sous un processus de bifurcation de type période-N est moins critique que celle de type Hopf. Ainsi, il est indispensable d'identifier précocement les bifurcations de type période-N pour améliorer l'efficacité d'usinage. Pour cela, nous avons développé une méthode d'identification du type et de la taille des bifurcations de période-N. Nous avons également prouvé l'efficacité des méthodes proposées, en utilisant deux modèles de processus de fraisage de référence. De plus, les méthodes proposées peuvent être utilisées pour le diagnostic de défaut d'autres systèmes dynamiques, tels que les systèmes de conversion d'énergie par modulation de largeur d'impulsion ou systèmes de paliers ou d'engrenage.

Mots clés : diagnostic cumulatif, détection précoce de broutage, bifurcations de période-*N*, identification, MSPR, niveau de risque, seuil.

Abstract

Cumulative diagnosis of dynamic systems requires the detection, identification, and characterization of incipient degradations. Its application to high-speed machining, for instance, could rely on period-N bifurcations phenomena analysis to detect and identify early-chatters and improve the quality of milling products and processes. Up to now, many efficient methods were proposed to detect early-chatter and identify period-N bifurcations. But these methods are struggling to implement these tasks reliably and accurately due to the complex nonlinear characteristics of their dynamic behaviors, the noise, and the variation of their operating conditions. The present thesis aims to develop and implement methods of early-chatter detection and period-N bifurcations identification within a real-time cumulative diagnosis approach. Aimed at early-chatter detection, we proposed three detection methods and one identification method for the cumulative diagnosis. The first method can be used to detect early-chatters remotely. The second one detects early-chatter quickly under specific operating and measuring conditions. However, in practice, the operating and measuring conditions are complex and variable. To adapt to different operating and measuring conditions, we proposed a third method, and the latter detects early-chatter reliably. It is also noted that in milling processes, the early-chatter can give rise to a bifurcation of period-N or Hopf type. The machining quality under the bifurcation process of the period-N type is less critical than that under the Hopf bifurcation type. To improve machining productivity and ensure the required machining quality, we can mill the workpiece under the condition of period-N bifurcations. Thus, it is compulsory to identify the early period-N bifurcations for improving machining productivity. For that purpose, we developed a method for identifying the type and size of the period-N bifurcations. We also proved the effectiveness of the proposed methods, using two benchmark milling process models. Besides, the proposed methods can be used for fault diagnosis of other dynamic systems, such as the pulse energy conversion systems or bearing or gearing systems.

Keywords: Cumulative diagnosis, Early-chatter detection, Period-*N* bifurcations, identification, MSPR, risk level, threshold.

Table of Contents

List of Publications

[1] Y. Zhao, K. H. Adjallah, A. Sava, Z. Wang, "Chatter Identification in Milling Process Based on Complementary Ensemble Empirical Mode Decomposition," ICMEM'2017, 11-12 December 2017, Metz, France. (Conférence sans actes)

[2] Y. Zhao, K. H. Adjallah, A. Sava, "Influence study of the intermittent wave amplitude vs. the sampling frequency ratio on ICEEMDAN mode mixing alleviation performance," 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19-21 November 2018, pp. 1-5, https://doi.org/10.1109/ICDSP.2018.8631700.

[3] Y. Zhao, K. H. Adjallah, A. Sava, Z. Wang, "Early chatter detection using MaxEnt and SPRT," CoDIT'2019 (6th International Conference on Control, Decision and Information Technologies), Paris, France, 23-26 April 2019, pp. 662-667, https://doi.org/10.1109/CoDIT.2019.8820670.

[4] Y. Zhao, K. H. Adjallah, A. Sava, Z. Wang, "Online incipient chatter detection based on once-per-revolution sampling and dynamic threshold variant," Proc. of IDAACS2019 (the 10th IEEE, International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications), Metz, France, 18-21 September 2019, pp. 1550-1555, https://doi.org/10.1109/IDAACS.2019.8924311.

[5] Y. Zhao, K.H. Adjallah, A. Sava, Z. Wang. "Period-*N* bifurcations identification in nonlinear dynamic systems based on simple moving average," to appear in the Proceedings of ICEASSM 2019, Sunyani, Ghana, 11-13 December 2019.

[6] Y. Zhao, K. H. Adjallah, A. Sava, Z. Wang, "MaxEnt feature-based reliability model method for real-time detection of early chatter in high-speed milling," in Press, ISA Transactions, July 2020, https://doi.org/10.1016/j.isatra.2020.07.022.

[7] Y. Zhao, K. H. Adjallah, A. Sava, Z. Wang, "A new framework of online early chatter detection in high-speed milling mechanical system and signal process," in preparation.

List of Figures

List of Tables

Chapitre 1 Présentation générale de la thèse

1.1 Motivation et portée du travail

L'industrie est un capital fondamental de tout système économique moderne. Dans l'industrie moderne, les systèmes dynamiques sont impliqués dans de nombreuses applications où ils assurent des fonctions clés. Les systèmes dynamiques incluent, entre autres, les diverses machines tournantes typiques (pompes, compresseurs, moteurs, générateurs, généralement des convertisseurs d'énergie électromécanique, …) ainsi que des convertisseurs de puissance électrique, etc.

Durant le fonctionnement, le comportement anormal des systèmes dynamiques dépend de leurs paramètres physiques et de réglage, de la dégradation des composants, des conditions de fonctionnement et de l'impact de l'environnement extérieur. Par exemple, la dégradation d'un roulement ou l'usure d'un outil de coupe peut provoquer des bruits lors du processus de fraisage [1]. L'apparition d'un comportement anormal affecte la fiabilité et les performances de fonctionnement des systèmes dynamiques, et porte atteinte à la sécurité des opérateurs (avec la survenue possible d'un accident grave) et à la qualité des produits.

Il est donc particulièrement important d'assurer pour les systèmes dynamiques, un niveau de fiabilité satisfaisant et de réduire les coûts de maintenance.

Le diagnostic des pannes participe donc efficacement à la stratégie pour assurer des conditions de fonctionnement sûres et optimiser l'utilisation des systèmes dynamiques. En général, les signaux recueillis à partir de systèmes dynamiques peuvent être utilisés pour identifier le mode de défaillance. En extrayant les informations de défaillance des signaux, il est possible de détecter et d'identifier les pannes. Ainsi, le diagnostic des pannes a été introduit pour assurer un fonctionnement sûr des systèmes dynamiques et permettre des décisions optimales en matière de maintenance.

En pratique, le défaut naissant est difficile à révéler dans un indicateur et est souvent enfouie dans un bruit d'arrière-plan. Il est donc difficile de le diagnostiquer de manière fiable. Pour diagnostiquer avec précision le défaut naissant, certains chercheurs ont proposé d'utiliser les méthodes de fusion de données. La fusion de données fait référence à l'intégration des informations provenant de plusieurs jeux de données pour des résultats plus consistants, fiables et utiles que ceux obtenus à partir d'un seul jeu de données [2]. Cette approche a été largement utilisée pour le diagnostic et le pronostic. Cependant, l'utilisation de plusieurs capteurs augmente la complexité des systèmes de surveillance et le coût de leur maintenance. Sans aucun doute, l'utilisation de plusieurs capteurs augmente la complexité du diagnostic, tout en diminuant la fiabilité du système de capteurs. Dans certains cas, l'installation de plusieurs capteurs pose des problèmes en raison de contraintes d'espace.

Par ailleurs, l'évolution du processus de dégradation des systèmes dynamiques est complexe. Les méthodes existante de fusion de données sont parfois limitées face au diagnostic de certains systèmes dynamiques.

Par exemple, un roulement peut intégrer une fonction « d'auto-réparation » [3][4]. Le processus « d'auto-réparation » implique un indicateur caractéristique de défaillance non monotone. Face à cette situation, il est difficile d'identifier de manière fiable le degré de dégradation de certains systèmes dynamiques en utilisant la fusion de données.

Pour pallier les inconvénients de la fusion de données, les chercheurs ont proposé l'approche cumulatif pour diagnostiquer de manière fiable un défaut naissant. L'approche cumulatif du diagnostic s'inspire de la théorie et du mécanisme du processus de dommages cumulatifs. Les dommages cumulatifs peuvent survenir en raison de nombreux facteurs, tels que le vieillissement, la fatigue, la dégradation, les chocs etc. Ainsi la probabilité de défaillance ou de panne augmente avec l'accumulation des dommages. L'approche de diagnostic cumulatif peut être classée en deux groupes : l'approche basée sur l'accumulation des dommages physiques et l'approche basée sur l'accumulation d'informations. L'approche basée sur l'accumulation de dommages physiques procède par la construction d'un modèle de dégradation basé sur la théorie et le mécanisme des processus de dommages cumulatifs [5][6]. Le modèle exponentiel et le modèle à risques proportionnels sont représentatifs des modèles de dégradation physique pour le diagnostic et le pronostic des défauts [7][8]. L'autre groupe d'approche de diagnostic cumulatif est basé sur l'accumulation d'informations. Cette approche se réfère au diagnostic des défauts en accumulant les informations successives observées à partir de données historiques et actuelles [7][9]. Elle diffère des méthodes de diagnostic traditionnelles qui dépendent des données actuelles observées et sans tenir compte des informations passées. Notre travail s'inscrit dans la deuxième approche de diagnostic, c'est-à-dire par l'accumulation d'informations historiques. Même si ce n'est pas spécifiquement indiqué, l'approche de diagnostic présentée dans ce qui suit est basée sur un processus accumulation d'informations.

Une fois que les données nouvellement observées sont disponibles, l'approche de diagnostic cumulatif met à jour les résultats préexistant de diagnostic en utilisant les nouvelles informations extraites des données nouvellement observées. Par exemple, les méthodes bayésiennes mettent à jour les résultats lorsque de nouvelles informations observées sont disponibles [7]. Puisque de nouvelles informations sont prises en compte, les résultats mis à jour sont améliorés. De cette façon, l'approche cumulative permettra de diagnostiquer les défauts précoces avec des résultats de plus en plus précis.

Le processus de diagnostic de défaillances comprend plusieurs tâches, telles que la détection des défauts, l'identification, la caractérisation, l'isolation et la localisation des défauts [10]. La détection des défauts est la première tâche du diagnostic. Elle vise à vérifier l'existence d'un défaut ou d'un dysfonctionnement dans un système. Pour atteindre cet objectif, la détection des défauts commence par l'acquisition/la surveillance/le contrôle des données, puis l'extraction des indicateurs caractéristiques du défaut, et enfin la prise de décision. Après avoir détecté le défaut, on doit identifier son type, sa forme et sa taille ou son intensité. Cette tâche est appelée identification des défauts et elle est vitale pour le diagnostic des défauts. L'isolement des pannes permet également de localiser le composant défectueux. Cette tâche affecte considérablement le temps et les

coûts de maintenance. Dans cette thèse, nous nous concentrons sur la détection, l'identification et la caractérisation des défauts.

Le processus de fraisage est choisi comme application. Il s'agit d'une approche d'usinage typique et largement utilisée en fabrication. L'un des phénomènes les plus courants rencontrés dans les processus de fraisage est le broutage dû à un phénomène de vibration auto-excité qui provoque une instabilité dans le processus de coupe. L'apparition du broutage réduit la productivité, accélère l'usure de l'outil de coupe et conduit à une faible qualité d'usinage. Par conséquent, le broutage précoce doit être détecté dès que possible pour éviter d'endommager la pièce et gagner du temps dans la boucle de contrôle du broutage. Actuellement, pour augmenter la vitesse d'enlèvement de matière, les paramètres d'usinage sont généralement fixés dans les régions de stabilité et près de la frontière d'instabilité. Dans ces conditions d'usinage, une légère variation des paramètres d'usinage peut conduire à un broutage léger à sévère. Pour certaines pièces particulières, une surface d'usinage très précise est nécessaire et un broutage, même léger, peut être intolérable [11]. Ainsi, la détection fiable du broutage naissant a une importance dans le processus de haute précision, en particulier pour les pièces de haute valeur. Par rapport au broutage sévère, nous qualifions également de précoce un broutage léger car il peut évoluer vers un broutage sévère suite à des changements des paramètres d'usinage. Sur la base de nos connaissances, de nombreuses recherches sur la détection du broutage sont proposées dans la littérature. Cependant, aucun ne peut détecter le broutage à un stade précoce avec une base théorique et en lien avec les paramètres de réglage de la machine. Pour combler cette lacune, nous avons proposé plusieurs méthodes de détection précoce du broutage basées sur l'approche de diagnostic cumulatif.

Un autre objectif de cette thèse est d'identifier rapidement et précisément les bifurcations de période-*N*. En effet, après avoir détecté les premiers signes d'un défaut ou d'un bavardage dans une approche de diagnostic, la phase suivante consiste à identifier la taille du défaut ou de la dégradation et de caractériser avec précision le mode de défaillance. Dans notre cas, le phénomène de bavardage est un processus dynamique périodique non linéaire dû aux vibrations. Pour son diagnostic, l'identification des composantes de fréquence harmoniques et sous-harmoniques dans les signaux de vibration est donc essentielle et peut être caractérisée par l'apparition de phénomènes de bifurcation. Pour atteindre cet objectif, nous proposons d'identifier les caractéristiques dynamiques intrinsèques des bifurcations de période-*N* liées au processus de bavardage. Une bifurcation de période-*N* est un phénomène non linéaire des systèmes dynamiques. En général, une bifurcation de période-*N* indique un comportement anormal. Par exemple, la bifurcation de période-*N* se manifeste souvent dans certains systèmes de conversion d'énergie électrique, lorsque leurs performances se dégradent. Il est donc nécessaire d'identifier les bifurcations de période-*N* pour assurer les performances élevées de fonctionnement de tels systèmes. Par ailleurs, dans certains systèmes dynamiques, la bifurcation de période-*N* est souhaitée pour implémenter certaines fonctions. Par exemple, la qualité d'usinage obtenue dans certaines conditions de bifurcation de période-*N* est meilleure que celle obtenue dans des conditions de bifurcation de type Hopf. Pour améliorer la productivité du processus de broyage, Honeycutt et Schmitz

[12][13] ont proposé de fraiser la pièce à usiner dans des conditions de bifurcations de période-*N*. Par ailleurs, une bifurcation en période-*N* peut révèler un comportement anormal dans certains systèmes dynamiques. Par exemple, la bifurcation de la période-*N* apparaît souvent dans certains systèmes de conversion d'énergie électrique lorsque leurs performances se dégradent ou qu'un défaut apparaît. Par conséquent, l'identification des bifurcations de période-*N* peut être utilisée pour identifier et caractériser des défauts de broutage. Jusqu'à présent, peu de travaux sur l'identification de la taille des bifurcations de la période-*N* sont rapportés. Afin d'identifier rapidement et précisément la taille des bifurcations de la période-*N*, cette thèse présente une nouvelle méthode d'identification des bifurcations de période-*N* dans une approche de diagnostic cumulatif.

1.2 Énoncé des problèmes et objectifs

L'approche de diagnostic cumulatif peut être appliquée à un vaste domaine, tel que la détection des défauts, l'identification des défauts et le pronostic. Dans cette thèse, nous utilisons l'approche de diagnostic cumulatif pour détecter les broutages précoces et identifier la bifurcation de la période-N. Un processus d'usinage a été sélectionné pour une étude de cas.

Dans un processus de fraisage à grande vitesse impliquant un bruit élevé, il est difficile de détecter rapidement et de manière fiable les broutages précoces. Dans le processus de fraisage à grande vitesse, la durée entre le début de broutage et le broutage mature est très courte. Elle peut être inférieure à 0,1 seconde [14]. Dans une période aussi courte, la plupart des stratégies de détection de broutage ne parviennent pas à extraire l'indicateur de broutage de manière fiable en raison du manque de sensibilité [15]. Ainsi, des méthodes d'extraction de l'indicateur de broutage à faible coût de calcul doivent être développées. De plus, au début du broutage, l'indication de broutage n'est pas apparente dans un environnement bruyant. Il est donc difficile de détecter de manière fiable le broutage précoces en utilisant des méthodes générales.

Les stratégies existantes peinent à mettre en œuvre rapidement et à identifier précisément les bifurcations de période-N en raison des caractéristiques non linéaires complexes de leurs comportements dynamiques et des conditions variables de travail. Par exemple, bien que la transformée de Fourier rapide (FFT) puisse identifier rapidement la bifurcation de la période-*N*, la résolution est faible lorsque la largeur de la fenêtre temporelle est étroite. Par contre une fenêtre de temps étendue peut augmenter la résolution, mais le délai de détection est également prolongé. En outre, la FFT peut ne pas identifier la taille de la bifurcation de période-N en présence de nombreuses composantes sous-harmoniques, harmoniques ou inter-harmoniques.

Pour aborder les deux problèmes mentionnés ci-dessus dans le processus de détection précoce du broutage et d'identification de la bifurcation de période-*N*, cette thèse propose trois méthodes de détection précoce de broutage et une méthode d'identification de période-*N* dans une approche de diagnostic cumulatif. Les méthodes de détection de broutage précoce proposées peuvent détecter les broutages précoces dans les processus de fraisage à grande vitesse en présence de bruits importants. La méthode proposée d'identification de la bifurcation de période-*N* identifie le type et la taille de la bifurcation de période-N de

manière rapide et précise.

Nous avons évalué les méthodes proposées en utilisant deux procédés de fraisage validés qui peuvent exprimer des propriétés non linéaires complexes.

1.3 Contributions de la thèse

Cette thèse comprend quatre contributions majeurs comme suit:

Détection précoce des broutages

(1) Nous avons proposé une première méthode de détection précoce du broutage basée sur le principe de l'entropie maximale (MaxEnt) et le test du rapport de probabilité séquentielle (SPRT), en utilisant la technique d'un échantillonnage par tour (OSPR) pour réduire la quantité de données échantillonnées. La méthode de détection précoce de broutage ainsi proposée, avec peu de données, est appropriée pour une implémentation avec un capteur sans fil pour la surveillance et la détection précoce de broutage à distance. Nous avons utilisé le test séquentiel du rapport de vraisemblance pour cumuler les informations de l'indicateur de broutage naissant. Les résultats obtenus grâce à cette contribution ont été publiés dans les actes de la conférence internationale CoDIT'2019.

Bien que la méthode proposée puisse détecter le broutage précoce, le broutage de période-2 associé à la bifurcation de période-2 n'est pas détecté de manière fiable. Pour le broutage quasi-périodique associé à la bifurcation de Hopf, chaque ensemble de données échantillonnées une fois par révolution a une fluctuation, qui peut être utilisée pour décrire l'apparence du broutage. Cependant, pour le broutge de période-2 associé à la bifurcation de période-2, il existe un ensemble de données échantillonnées une fois par tour sans fluctuation [16]. Il est très difficile de distinguer le signal sans broutage de celui avec broutage de période-2 sur la base des données échantillonnées une fois par tour. Ainsi, la méthode proposée basée sur OSPR et SPRT doit être développée pour des applications pratiques.

(2) Nous avons proposé une deuxième méthode de détection rapide du broutage précoce basée sur l'échantillonnage multiple par révolution (MSPR) et la méthode de fiabilité du second ordre (SORM). Le MSPR surmonte les limitations dues à l'OSPR de la première méthode présentée plus tôt. L'indicateur MaxEnt obtenu en utilisant la technique d'échantillonnage SORM résout le problème de détection en temps réel du broutage précoce tout en garantissant sa fiabilité. Les résultats obtenus par simulation numérique expérimentation confirment la capacité de cette méthode à détecter rapidement les broutages précoces. Les résultats obtenus grâce à cette contribution ont été publiés dans le journal ISA Transactions.

(3) Enfin, nous avons proposé une troisième méthode de détection fiable précoce du broutage basée sur la l'échantillonnage technique MSPR avec des seuils adaptatifs basée sur deux niveaux de risque de décision. Par rapport à la technique OSPR proposée dans [17], l'échantillonnage multiple par tour réduit la largeur de la fenêtre temporelle ainsi que le délai de détection. En outre, l'utilisation de seuils

adaptatifs basés sur deux niveaux de risque de décision améliore la fiabilité de la détection, réduit les limitations du seuil liées aux principe des petites probabilités, et est bien adaptée à des mesures sous influence de bruits significatifs. Les résultats obtenus avec cette méthode seront soumis à une revue internationale pour publication.

Identification de la bifurcation de la période-N

(4) Dans cette partie, nous avons d'abord appliqué une décomposition synchrone au signal de bifurcation période-*N*. Ensuite nous avons développé une méthode d'identification rapide et plus précise des bifurcations de période-*N*, basée sur la moyenne mobile simple (SMA) et la décomposition synchrone. La méthode comprend deux étapes: l'identification du type et de la taille de la bifurcation de période-*N*. Nous avons alors utilisé la moyenne mobile simple pour déterminer le type de bifurcation période-*N* en repoussant la limitation de l'échantillonnage sous-harmonique. Pour identifier la taille de la bifurcation de période-*N*, nous proposons d'utiliser la décomposition synchrone dont nous avons donné une preuve mathématique simple. Les résultats obtenus avec cette méthode seront également soumis à une revue internationale pour publication.

1.4 Organisation de la thèse

La thèse comprend six chapitres agencés comme dans la Figure 1.1.

Le chapitre 1 présente d'abord la motivation et la portée de la thèse, puis le problème et enfin les objectifs, tout en soulignant les apports du travail proposé.

Le chapitre 2 passe brièvement en revue l'état de l'art sur l'approche de diagnostic cumulatif, la détection du broutage naissant et l'identification des bifurcations de période-*N*. L'état de l'art montre que l'approche cumulative de diagnostic est pertinent pour détecter le broutage naissant et identifier les bifurcations de période-*N*.

Le chapitre 3 propose et présente en détail trois méthodes de détection de broutages naissants. Les méthodes reposent sur des méthodes d'échantillonnage spécifiques, d'extraction d'un indicateur caractéristique du défaut, et de décision appropriée.

Le chapitre 4 propose une méthode d'identification de bifurcations des période N basée sur le SMA et la décomposition synchrone. La méthode proposée comprend deux étapes : l'identification du type de bifurcation, et l'identification de la taille de la bifurcation de période-*N*. Enfin, nous avons fourni une preuve de consistance de la décomposition synchrone.

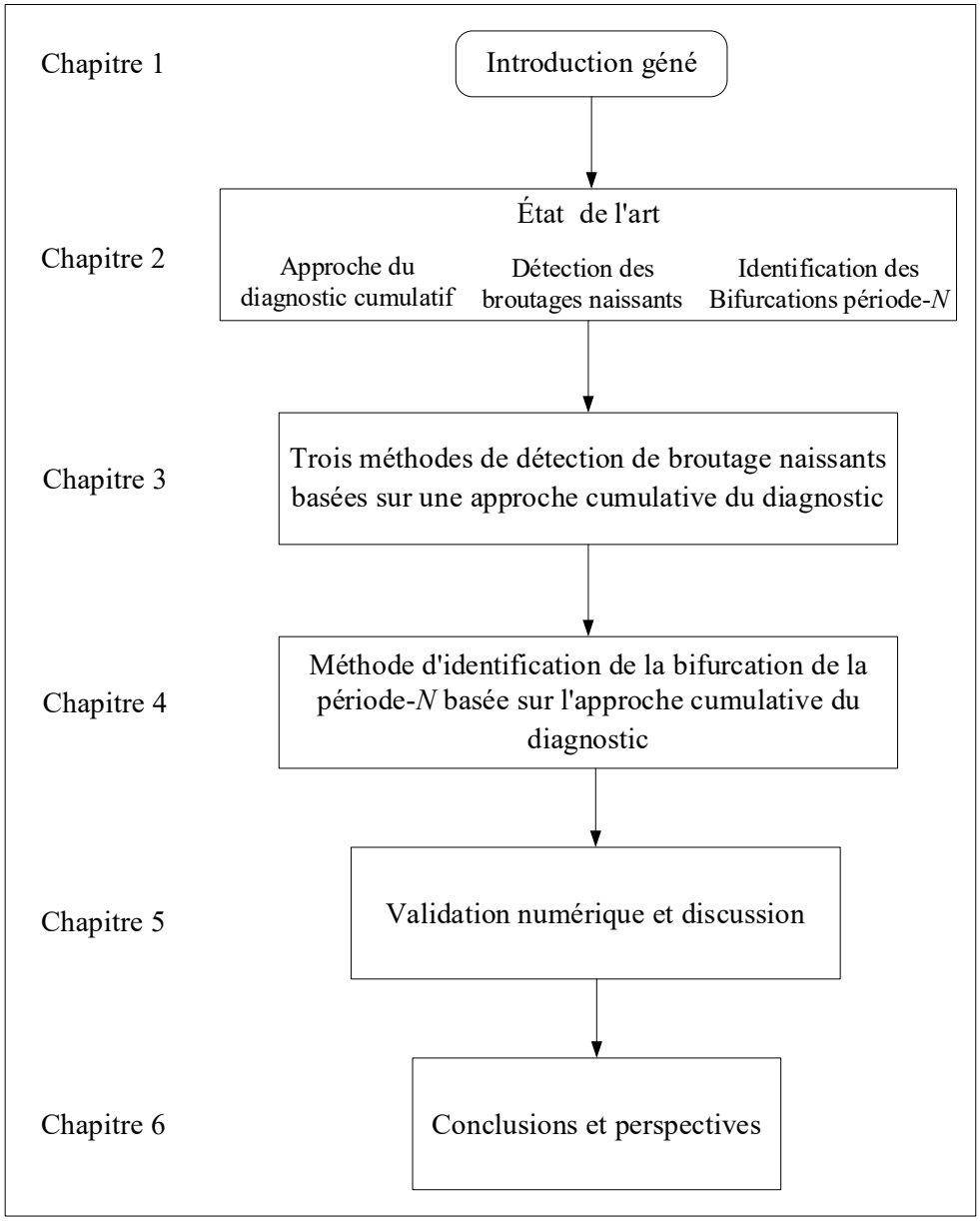


Figure 1.1 Organisation de la thèse.

Le chapitre 5 est dédié à l'évaluation, sur la base de deux modèles de fraisage, des trois méthodes proposées pour la détecter les broutages naissants, ainsi que de la méthode d'identification des bifurcations de période-*N*.

Enfin, le chapitre 6 scelle la thèse, avec quelques conclusions et perspectives.

Chapter 1 General introduction

1.1 Motivation and scope of work

The industry is a core asset of any modern economic system. In the modern industry, dynamic systems are involved in many applications where they perform key functions. Dynamic systems include, among others, various rotating machines (pumps, compressors, motors, generators, typically electromechanical power drives, …), as well as electrical power converters, etc.

During the operation process, dynamic systems may behave abnormally depending on their physical and setting parameters, degradation of parts, operation conditions, and influences of the external environment. For instance, the degradation of a bearing or wearing of a cutting tool can cause chatter in the milling process [1]. The appearance of abnormal behavior affects the reliability and operating performance of dynamic systems and undermines operators' security (with the possible occurrence of a severe accident) and product quality. Therefore, it is particularly crucial to ensure dynamic systems a satisfactory level of reliability and to reduce maintenance costs.

Fault diagnosis is a practical part of the strategy to assure safe operating conditions and optimize the dynamic systems' usage. In general, the signals collected from dynamic systems can be used to identify the failure mode. It is possible to detect and identify the failure by extracting related information from the signals. Thus, fault diagnosis has been introduced to assure a safe operation of dynamic systems and allow optimal maintenance decisions.

In practice, the incipient fault is difficult to reveal through a feature and is often buried in background noise, making it difficult to diagnose reliably. To diagnose the incipient fault accurately, some researchers rely on data-driven approaches based on statistical data processing, signal processing, sensors data fusion, etc. Data fusion refers to integrating the information from multiple datasets to obtain more consistent, reliable, and useful results than from an individual dataset [2]. This approach has been used for diagnosis and prognosis widely. However, the use of multiple sensors increases the complexity of the surveillance systems and their maintenance cost. Undoubtedly, the use of multiple sensors increases the diagnosis complexity while decreasing the sensor system reliability. In some cases, the installation of multiple sensors is difficult due to space constraints.

Besides, the evolution of the degradation process of dynamic systems is complex. The existing data fusion sometimes fails to diagnose some dynamic systems. For instance, a bearing may incorporate a function of

“self-repair.[3][4]” The process of "self-repair" involves a non-monotonic degradation feature. In this situation, it is difficult to identify reliably the degradation degree of some dynamic systems using data fusion.

Researchers proposed the cumulative approach to diagnosing an incipient fault reliably to overcome the disadvantage of data fusion. The cumulative approach of the diagnosis is inspired by the cumulative damage process theory and mechanism. The cumulative damage can arise due to many factors, such as aging, fatigue, degradation, shocks, etc. Thus, the fault or failure probability increases with the accumulation of damages. The cumulative diagnosis approach can be categorized into two groups: the physical damage accumulation-based approach and the information accumulation-based approach. The physical damage accumulation-based approach proceeds by building the physical degradation models [5][6]. The exponential model and the proportional hazards model are representative of physical degradation models for fault diagnosis and prognosis [7][8]. The other group of cumulative diagnosis approach is based on information accumulation. This approach refers to fault diagnosis by accumulating historical and current observed data [7][9]. It differs from the traditional diagnosis methods, which depend on the current observed data without considering past information.

Our work belongs to the second group of cumulative diagnosis approach, i.e., by accumulating historical data and information. Although not specifically stated, the diagnosis methods introduced in the following are based on information accumulation. Once newly observed data become available, the cumulative diagnosis method can update the pre-existing diagnosis results using the new information extracted from the newly observed data. For instance, the Bayesian-based methods update the results when the new observed information becomes available [7]. As new information is taken into account, the updated results are improved. This way, the cumulative approach will allow diagnosing early faults with increasingly accurate results.

The fault diagnosis process includes several tasks, such as detection, identification, characterization, isolation, and location [10]. Fault detection is the first task of the fault diagnosis. It aims to check the existence of a fault or malfunction in a system. For this purpose, fault detection starts with the data acquisition/surveillance/monitoring, then feature extraction, and lastly, decision making. After detecting the fault, one must identify its type, shape, size, or intensity. This task is called fault identification and is vital for fault diagnosis. Fault isolation also allows for locating the defective component. This task affects the maintenance time and costs significantly. In this thesis, we focus on the fault detection, identification, and characterization of faults.

The milling process is chosen as an application. The milling process is a typical machining approach and widely encountered in the manufactory. One of the most common phenomena to be handled in the milling process is the chatter, which is a self-excited vibration phenomenon causing instability in the cutting process. The chatter appearance reduces productivity, accelerates wear of the cutting tool, and leads to low

machining quality. Therefore, the early-chatter should be detected as soon as possible to avoid the work-piece damage and save time in the chatter control loop. At present, to increase the material removal rate, the machining parameters are usually set within the stability regions and nearby the instability boundary. Under such a machining condition, a slight variation of machining parameters may lead to a slight or severe chatter. A high precision machining surface is necessary for some particular parts, and even slight chatter may be intolerable [11]. Thus, the reliable slight chatter detection has significance in the high precision milling process, especially for valuable parts. Compared to the severe chatter, we also call the slight chatter the early-chatter because it can evolve toward the severe chatter with the machining parameters changes. Based on our knowledge, many research works on chatter detection are proposed in the literature. However, none can detect chatter at an early-stage with a theoretical basis and dependence on machine parameters. To fill this gap, we proposed several methods for early-chatter detection based on the cumulative diagnosis approach.

Another objective of this thesis is to identify the period-*N* bifurcations quickly and accurately. Indeed, after detecting the first signs of a fault or chatter in a diagnostic approach, the next phase is to identify the fault or degradation size and characterize the failure mode accurately. In our case, the chatter phenomenon is a nonlinear periodic dynamic process due to vibrations. For its diagnosis, the identification of harmonic and subharmonic frequency components in the vibration signals is therefore essential and can be characterized by the appearance of bifurcation phenomena. To achieve this objective, we propose to identify the intrinsic dynamic characteristics of the period-*N* bifurcations related to the chatter process. The period-*N* bifurcations are nonlinear phenomena of dynamic systems. In some dynamic systems, the period-*N* bifurcations are desired for implementing specific functions. For instance, the machining quality obtained under certain conditions of period-*N* bifurcations is better than that obtained under conditions of type Hopf bifurcation. To improve the milling process productivity, Honeycutt and Schmitz [12][13] proposed to mill the workpiece under conditions of the period-*N* bifurcations. Besides, a period-*N* bifurcation can reveal an abnormal behavior in some dynamic systems. For instance, the period-*N* bifurcation often looms up in some electrical power conversion systems when their performance degrades, or a fault appears. Consequently, period-*N* bifurcations analysis can be used to identify and characterize chatter faults. Up to now, few research works on size identification of period-*N* bifurcations are reported. In order to identify the size of period-*N* bifurcations quickly and accurately, this thesis presents a new period-*N* bifurcations identification method within the cumulative diagnosis approach.

1.2 Problems statement and objectives

The cumulative diagnosis approach can be applied to a wide area, such as fault detection, fault identification, and prognosis. In this thesis, we employ the cumulative diagnosis approach to detect early-chatter and identify period-*N* bifurcations. A milling process was selected for a case study.

In a high-speed milling process involving heavy noise, it is challenging to detect early-chatter quickly and reliably. In the high-speed milling process, the duration from chatter onset to mature chatter is short. It can be less than $0.1s$ [14]. In such a short period, most chatter detection strategies fail to reliably extract the chatter feature due to the lack of sensitivity [15]. Thus, extraction methods of chatter feature with low computational cost need to be developed. Also, at the early stage of the chatter, the chatter feature is not apparent and can be emerged in a noisy environment. It is, therefore, challenging to detect early-chatter reliably using general methods.

The existing strategies struggle to quickly and accurately identify period-N bifurcations due to the complex nonlinear characteristics of their dynamic behaviors and the variational working condition. For instance, although fast Fourier transform (FFT) can quickly identify period-N bifurcations, the resolution is low when the time-window width is narrow. Conversely, the extended time-window can increase the resolution, but the detection delay is also extended. The FFT may also fail to identify the period-N bifurcations size when there are many sub-harmonics, harmonics, or inter-harmonics components.

For addressing the abovementioned two issues in the process of early-chatter detection and period-N bifurcation identification, this thesis proposes three early-chatter detection methods and one period-N identification method based on the cumulative diagnosis approach. The proposed early-chatter detection methods can detect early-chatter in high-speed milling processes involving heavy noise. The proposed period-N bifurcation identification method identifies the type and size of the period-N bifurcation quickly and accurately.

We evaluated the proposed methods using two validated milling processes that can express complex nonlinear properties.

1.3 Thesis contributions

This thesis includes four contributions, as follows:

Early-chatter detection

(1) We proposed the first method of early-chatter detection based on the maximum entropy (MaxEnt) principle and the sequential probability ratio test (SPRT) while using the technique of once sampling per revolution (OSPR) to reduce the amount of sampled data. With a few data, the proposed early-chatter detection method is suitable for remote monitoring and early-chatter detection. We used the sequential probability ratio test to accumulate the incipient chatter feature information. The results obtained through this contribution were published in the CoDIT'2019 international conference.

Although the proposed method can detect early-chatter, the period-2 chatter associated with the period-2 bifurcation is not detected reliably. For the quasi-periodic chatter related to the Hopf bifurcation, each set of OSPR data has fluctuation, which can be used to depict the chatter's appearance. However, for the period-2 chatter associated with the period-2 bifurcation, there is a set of OSPR data without the

fluctuation [16]. It is very hard to distinguish the chatter-free signal from the period-2 chatter, based on that set of OSPR data. The proposed method, based on OSPR and SPRT, needs therefore further development for practical applications.

(2) We proposed a second method for quick early-chatter detection using the multiple sampling per revolution (MSPR) and the second-order reliability method (SORM). The MSPR overcomes the limitation of OSPR introduced in the last paragraph. The proposed SORM with the MaxEnt feature solves the issue of real-time early-chatter detection while ensuring its reliability. The results obtained by the numerical simulation experiment confirmed that our method could detect early-chatter quickly. The results obtained through this contribution were published in the journal ISA Transactions in 2020.

(3) Lastly, we proposed a reliable early-chatter detection method based on the MSPR technique, with an adaptive threshold based on two decision risk levels. To obtain reliable detection results, we proposed to detect early-chatter based on MSPR and adaptive threshold based on two decision risk levels. Compared to the proposed OSPR technique in [17], the MSPR technique reduces the time-window width as well as the detection delay. Besides, using the adaptive threshold based on two decision risk levels improves the detection reliability, reduces the limitations of the threshold based on only one decision risk level, because of the small probabilities principle, with better immunity to noises. The results obtained through this method will be submitted to an international journal for publication.

Period-N bifurcations identification

(4) In this part, first, we applied the synchronous decomposition to the period-*N* bifurcation signal. Then we developed a quick and more accurate period-*N* bifurcations identification method based on the simple moving average (SMA) and synchronous decomposition. The proposed method includes two steps: type identification of the period-*N* bifurcations and size identification of period-*N* bifurcations. We then used the SMA to identify the type of period-*N* bifurcations and overcome the limitations of sub-harmonic sampling. To identify the size of period-*N* bifurcations, we propose using the synchronous decomposition, of which we provided a simple mathematical proof. The results obtained with this method will also be submitted soon to an international journal for publication.

1.4 Organization of the thesis

The thesis includes six chapters laid as in Figure 1.1 below.

Chapter 1 introduces the thesis' motivation and scope first, then the problem, the objectives, and lastly, highlights the contributions of the proposed work.

Chapter 2 reviews state of the art on the cumulative diagnosis approach, early-chatter detection, and period-*N* bifurcations identification briefly. State of the art points out that the cumulative diagnosis approach can effectively detect early-chatter and identify period-*N* bifurcations.

Chapter 3 proposes and details three early-chatter detection methods. The methods rely on specific

sampling techniques, extraction of the fault feature, and appropriate decision-making.

Chapter 4 proposes a period-*N* bifurcations identification method based on the SMA and the synchronous decomposition. The proposed period-*N* bifurcations identification method includes two steps: the period-*N* bifurcations type identification and the period-*N* bifurcation size identification. Lastly, we provided proof of the consistency of the synchronous decomposition.

Chapter 5 is dedicated to evaluating, through two milling models, the performance of the three methods proposed for early-chatter detection and the period-*N* bifurcations identification method.

Finally, Chapter 6 seals the thesis, with some conclusions and prospects.

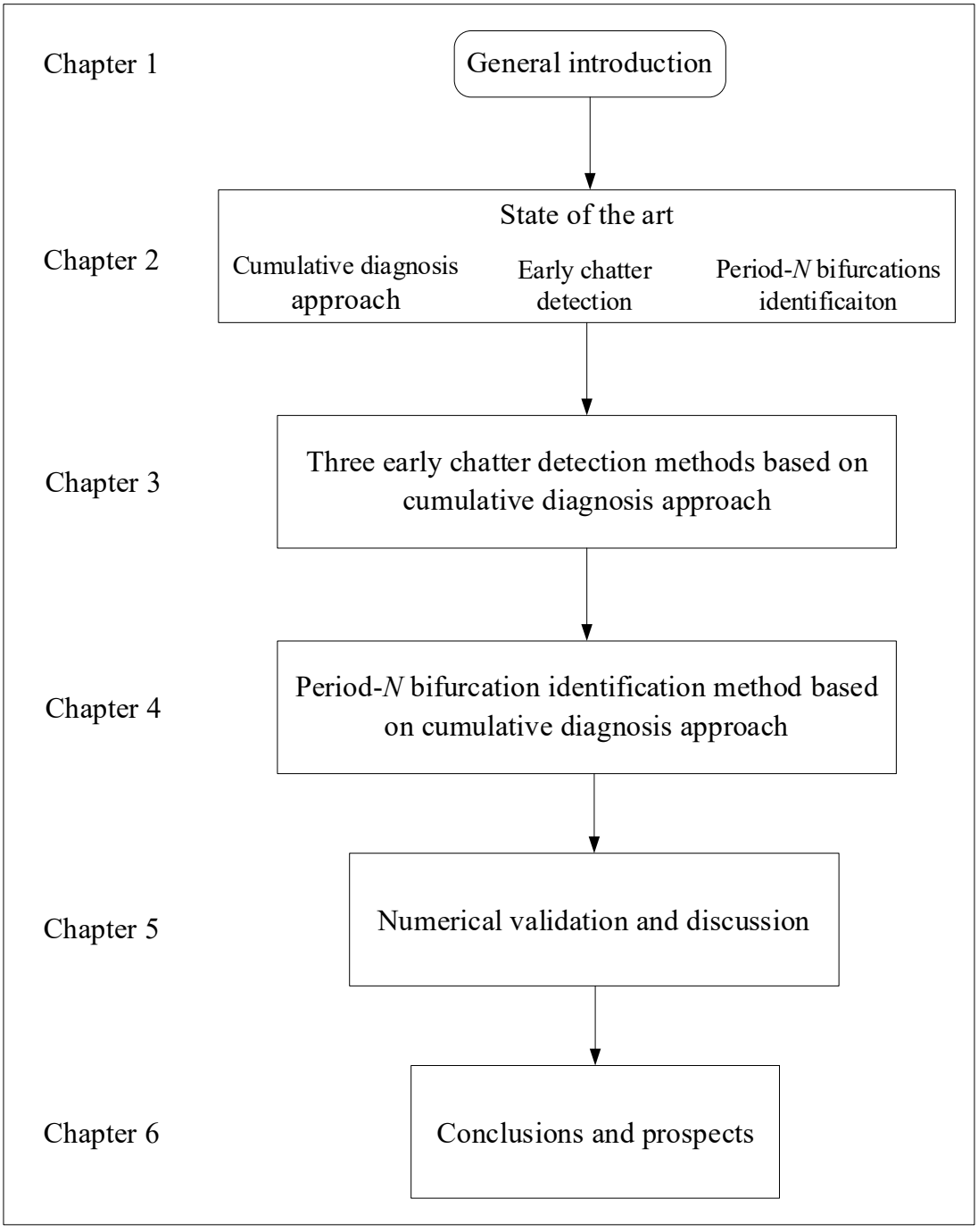


Figure 1. 1 Organization of the thesis.

Chapter 2 State of the art

2.1 Introduction

Cumulative diagnosis in dynamic systems aims to early degradation detection. As introduced in the last chapter, the cumulative diagnosis approach is inspired by the cumulative damage process theory and mechanism. By accumulating the inapparent information in historical data, reliable diagnosis results can be obtained. Up to now, many cumulative diagnosis methods were presented in broad fields of applications.

The cumulative diagnosis approach can be used for early-chatter detection and period-*N* bifurcations identification. In the machining field, the chatter can cause instability, reduce productivity, accelerate wear of the cutting tool, and lead to low machining quality [18]. For the purposes of avoiding the work-piece damage and saving time in the chatter control loop, the early-chatter should be detected as soon as possible [19]. At the early-chatter stage, the chatter feature is inapparent and cannot indicate the chatter reliably. The cumulative diagnosis approach can obtain reliable detection results by accumulating the inapparent chatter feature.

In all dynamic systems, operating at a given type of bifurcation exhibits similar characteristics. However, systems operating at a different type of bifurcation express a different character. Generally, the presence of the period-*N* bifurcations refers to a system fault or function fault in a mechanical and electrical system. Therefore, by identifying the type and size of the period-*N* bifurcation, one may determine the fault type and size. Besides, the period-*N* bifurcations identification may be applied to improving or predicting the performance of the dynamic system [12][19]. For instance, Honeycutt and Schmitz [12][13] proposed to improve the milling process productivity using the machining parameters, which can lead to the period-*N* bifurcations. Therefore, period-*N* bifurcations identification has great significance in practice.

To develop some early-chatter detection methods and period-*N* bifurcations identification method for implementing the cumulative diagnosis approach, we provide state of the art on cumulative diagnosis methods, early-chatter detection, and period-*N* bifurcations identification in this chapter, respectively. The early-chatter detection and period-*N* bifurcations identification can be seen as two applications of the cumulative diagnosis approach.

We organize the rest of the chapter as follows. Section 2.2 reviews cumulative diagnosis methods according

to their functions. Furthermore, as a case study, we focus our attention on early-chatter detection and period-N bifurcations identification. Sections 2.3 and 2.4, respectively, review relevant research work in the literature. Also, we point out the potential of the cumulative diagnosis approach for these two topics briefly. Finally, Section 2.5 concludes this chapter.

2.2 State of the art on cumulative diagnosis methods

In early degradation monitoring and detection, there exists a challenge. The noise and some components in the signal, collected in order to monitor the current state of systems, affect the magnitude of inapparent fault features. Therefore, it is not a trivial task to detect incipient degradation quickly and reliably. In this context, the cumulative diagnosis approach offers the following four advantages to cope with the cumulative character of damages underlying all continuous degradation processes: (1) reducing the random noise residing in the signal or incipient fault feature; (2) amplifying the fault symptom in the feature; (3) filtering or reducing some frequency components residing in the signal; (4) consolidating the degradation information. The following section will review the cumulative diagnosis methods from these four aspects.

（1） Reducing the random noise residing in the signal or estimated parameters.

Based on the cumulative diagnosis approach, the cumulative diagnosis methods for reducing random noise can be classified into two categories: cumulative average based methods, and Bayesian-based methods.

The cumulative average based methods refer to calculating the sum of the current data and all or part of available observed historical data. Dividing the result by the total number of data involved yields the average. The summation or averaging of the signals or the estimated parameters reduces the random noise in the data.

The moving average (MA) and time-synchronous average (TSA) algorithms are representative of the cumulative average based methods [20][21]. The MA calculates the average value of samples with a specific time-window (width) recursively. In comparison, the TSA calculates a group of average values recursively according to the specific period of the signal.

The MA is a finite impulse response filter that calculates weighted average values from the data recursively. The MA has two critical parameters: the time-window width and the data weights at different positions inside the time-window. For a given series of data and a preset the time-window width and data weights, the MA enables filtering the high-frequency noise and highlighting the longer-term trends or cycles of the signal evolution. The small fault feature merged in the random noise can be highlighted, and then the incipient fault can be detected [22].

The time-window width is an essential parameter with a high impact on the filtering performance of the MA. A more extended time-window width allows better performance for filtering the random noise. However, these results in more significant distortions will be produced. Nowadays, it does not exist the standard criteria for setting the time-window width effectively. In practice, the time-window width is

generally set according to a trade-off between filtering the performance and reducing the distortion. In power systems, the optimal width of the MA time-window is estimated using the Amari index [23]. The width of the MA time-window with the minimum Amari index is chosen as the optimal width. The optimal width of the MA time-window is determined to separate the fast and slow fluctuations based on using a threshold. On the forecasting basis, the optimal width of the MA time-window is determined according to the size and frequency of structural changes. A wider time-window is better when the conditions do not vary, whereas an optimal time-window width should be one chosen after the appearance of a severe structural change [24].

The data weights also impact the filtering performance of the MA. The significance of the historical and current data determines their weights. According to data weights, one can distinguish two types of MA methods: the simple moving average (SMA) and the weighted moving average (WMA).

The SMA is the simplest MA method since the data weights are equal. This method is used in situations where the data have the same or almost the same significance. When the time-window width is N, the data weights are set equal to $1/N$. For the SMA, one can see that it has only one parameter, i.e., the time-window width. As mentioned above, setting the time-window width is based on the application generally. A more extended time-window width can be set if the expected longer-term trends or cycles have no frequency or a low frequency. In fault diagnosis, the SMA is employed frequently at the signal preprocessing or processing stages for filtering purposes. Wang and Tse [25] proposed to assess the degradation tendency of the slurry pump impeller by using the SMA at the signal processing stage. The results depend both on the current measurements and the four previous ones. Therefore, the degradation tendency was highlighted. Attuati et al. [26] used the SMA to reduce the random noise in the residuals produced by a model-based fault diagnosis method. Since less random noise is in the residuals, the drawback of the restrictive trade-off between the false-alarm and the missed-detection rates was reduced.

To detect the slowly varying incipient faults, the SMA is also combined with the principal component analysis (PCA) and designated component analysis (DCA) for developing the average accumulative (AA) based PCA and DCA [9][27]. Aimed at some sets of sampled multivariate series $Y(i, j)$, the SMA is used to reduce the noise by calculating the average first. Then, the PCA projects the average to principal components, and the DCA projects the average to designated patterns. Due to the filtering property of the SMA, the obtained principal components or designated patterns have a higher signal to noise ratio (SNR) than that obtained from the original PCA or DCA [27].

In practice, the current health state of the machine is more correlated with the recent measurements. Considering this point, researchers proposed to use the weighted moving average (WMA) that gives a higher weight on the more recent measurements. For the WMA, the challenge is how to set the data weights. The WMA can obtain a better performance than the SMA if the weights are appropriately set.

The commonly used WMA is the exponential moving average (EMA) [28], also known as an exponentially weighted moving average (EWMA). The data weights of EMA are set to decrease exponentially for each historical datum. For the distant past data, the weights are small but never reach zero. The EMA can smooth the measured signal. In a gas path diagnostic, the smoothing effect on the overall trend shifts was produced by using the EMA [29].

The data can also be accumulated one time-window by one time-window synchronously. It is another way of averaging, according to the periodicity of the signal. The TSA refers to calculate the average of the synchronous data according to the time-window width. In the TSA, the time-window width is set corresponding to the desired synchronous frequency.

Due to the synchronous average of the signal, the random noise and non-synchronous frequency are reduced or removed. Thus, the expected signal components can be highlighted and extracted [21]. Given the signal $s(t)$, the TSA can be expressed as follows:

$$\bar{s}(t) = \frac{1}{N}\sum_{n=0}^{N-1} s(t+nT), \qquad 0 \le t < T, \tag{2-1}$$

where T represents the time-window width, also called the averaging period. N is the sample number for the accumulation. For the details about TSA, see the references [30] [31]. Up to now, the TSA has been a popular time-domain signal analysis method and used widely for fault diagnosis in the electrical and rotating machinery [30][21][32].

It should be noted that the TSA needs synchronous data. The tachometer, magnetic encoder, and optical encoder are used to implement synchronous sampling. For the purpose of obtaining the synchronous data, these sensors need to be installed in the systems. However, for a dynamic system that does not have enough space to install the sensors, it is difficult to obtain synchronous data. Aimed at this problem, some researchers proposed to resample the sampled data according to the estimated fundamental frequency of the signal (e.g., the rotating frequency for a rotating machine) [33].

Besides the property of the random noise reduction, the SMA and TSA enable filtering specific frequencies. We will review their performance of filtering specific frequencies in the sequel.

The cumulative average based methods also refer to noise filtering. In general, a filter processes an input signal x and produces a filtered output signal y. The different types of filters feed a memory with different amounts of historical input or output data for their implementations. The cumulative average based methods abovementioned consider only the historical input data of the filter. Besides the historical input data, the historical output data of the filter can also be considered for developing the cumulative average based methods. Based on this approach, researchers presented some cumulative methods, such as the autoregressive (AR) filter and the autoregressive moving average (ARMA) filter [34].

The cumulative average based methods can filter random noise effectively and quickly. In the cumulative average based methods, the main procedure is the averaging process. Therefore, the computational cost of

the cumulative average based methods is cheap. For the random noise, the averaging procedure has a high filtering performance. However, it is a challenge to determine the parameters of the cumulative average based methods. For instance, it is difficult to determine the optimal time-window width and data weights for the MA. The issue of the parameters determination of the cumulative average based methods may be addressed thoroughly or partly if we understand the signal model.

The second class of cumulative diagnosis methods that triggered our attention is represented by the Bayesian-based methods. This class of cumulative diagnosis methods is based on the Bayes theory, which provides a diagnosis result by combining different prior information and the information (or evidence) contained in the current data [35]. In the Bayesian paradigm, the result is a probability that measures a belief degree of a fault origin. Bayesian-based methods aim to update the prior belief degrees to the posterior belief degrees depending on the obtained new evidence. On the one hand, the updated posterior belief can support a better decision than the prior one due to considering the new evidence. On the other hand, a decision-making process based on the updated posterior belief is better than that based on only the new evidence since the prior belief is incorporated to improve the decision accuracy [36].

The Bayesian-based methods are used for fault diagnosis by accumulating the obtained evidence. In the literature, some cumulative fault diagnosis methods are reported based on the Bayesian-based methods.

In fault diagnosis, the well-known Bayesian-based methods include several algorithms. These algorithms can be categorized into several topics, such as Bayesian estimation and Bayesian testing [36].

The Bayesian estimation refers to estimate the unknown parameter θ or the unknown state of the dynamic systems. In the Bayesian estimation, the estimated parameter or state of dynamic systems can minimize the expected loss. In other words, the estimated parameter allows losing the least amount of accuracy.

In the fault diagnosis based on Bayesian estimation, the estimated parameters or state represent the system health-state and can be used to diagnose faults. The estimated parameters could be the system parameters, such as stiffness or damping of a dynamic system, or model parameters in general [37][38], such as Markov model parameters or exponential model parameters. In the mechanical or electronic systems, the system or model parameters represent their health conditions. The faults can be diagnosed directly when the estimated parameters cross the preset thresholds. After the appearance of a fault, the behavior of the dynamic system becomes abnormal. The fault feature can be extracted from the abnormal behavior for diagnosing the faults. Therefore, the fault diagnosis can be implemented by estimating the system or model parameters.

To estimate the parameters or the state of the dynamic system, researchers proposed many methods. The Kalman filter and particle filter are two representative Bayesian estimation methods in the fault diagnosis [39]. Kalman filter is to estimate the parameters or the state of the linear and Gaussian dynamic systems. The Kalman filter is an unbiased minimum variance estimator and can obtain an exact solution for the linear and Gaussian dynamic systems. Ondel et al. [40] proposed to use the Kalman filter to estimate and predict unknown states for detecting failure appearance. In the case of insufficient data, the Kalman filter can

decrease the number of measures and predict a linear dynamic system's unknown states. Conversely, for the nonlinear and non-gaussian dynamic systems, the performance of the Kalman filter is limited and unsatisfactory.

Aimed at the filtering problem of the nonlinear and non-gaussian dynamic system, researchers presented some nonlinear filters. The particle filter is commonly employed to estimate the parameters or the state of the nonlinear dynamic systems for fault diagnosis [41,42]. Rigatos [43] compared the performance of the Kalman filter and the particle filter for estimating the state of a DC motor. The particle filter has a better performance in estimating the state than the Kalman filter because the particle filter is not based on the assumption of Gaussian noise and linearization. However, the particle filter needs a large number of particles to achieve a highly accurate estimation. A large number of particles leads to a high computational cost. Therefore, the particle filter is not suitable for the real-time fault diagnosis.

The Bayesian testing is to test the hypothesis depending on the prior information and the available data. According to the Bayesian theory, the posterior probabilities of hypotheses can be calculated for making decisions. The maximum a posteriori (MAP) estimation is commonly employed for obtaining all hypotheses probabilities in decision-making. We can compare all hypotheses probabilities and then accept the hypothesis with the highest posterior probability. One can declare that the probability error is minimized since the posterior probability of the accepted hypothesis is the highest. Therefore, Bayesian testing is appropriate to conduct the hypothesis test [44]. Bayesian testing can obtain better results than the frequentist methods if the prior information is set appropriately.

In fault diagnosis based on Bayesian testing, some decision-making methods are more reliable. There are two actions in the simple Bayesian testing, i.e., acceptation and rejection, and two decision risk levels. The null hypothesis may be rejected when it is true (corresponding to the first decision risk level). The null hypothesis may be accepted when it is not true (called second decision risk level). To consider these two decision risk levels, the sequential probability ratio test (SPRT) [45], a Bayesian testing algorithm, is proposed [46]. Under the condition of specifying two decision risk levels, the SPRT can make decisions with satisfactory results.

In the SPRT, the log-likelihood ratio (LLR), defined as the logarithm of the likelihood ratio of the above two hypotheses, is accumulated from the first measurement to the current measurement [47]. According to the cumulative LLR and preset upper and low thresholds, three decisions can be made. The process will be seen as a normal (abnormal) process if the cumulative LLR is greater (less) than the preset upper (low) threshold. Otherwise, the test needs to continue until the preset upper or low threshold is crossed. Comparing the famous Shewhart control chart, the SPRT considers the information of all measurements. The new information of the likelihood ratio is accumulated sequentially. Therefore, the unapparent fault feature is accumulated to be an apparent index for indicating the fault appearance. Thus, the incipient fault can be detected easily by using powerful SPRT. Nowadays, the SPRT has been employed widely for fault

diagnosis in a wide range of fields, such as medicine [48], sewer network [49], chemical industry [50], and manufacturing industry [17]. To detect early-chatter reliably in the milling process, Zhao et al. [17] proposed to use the SPRT to reduce the decision risk level of chatter detection. Also, the SPRT suits the threshold following a distribution rather than a constant [51].

The Bayesian-based methods consider prior knowledge and accumulate the information from the newly obtained data. Therefore, the fault diagnosis methods based on the Bayesian-based methods yield more accurate results than other methods that consider only information from the newly obtained data. For the soft fault, such as the system degradation, the recently obtained data include the most relevant information on the new system state. However, the historical data also include some fault information of the system. The diagnosis result can be more accurate if the historical data are also considered in the diagnosis process. The Bayesian methods provide a framework to combine the information embedded in the recently obtained data for updating the knowledge on the state of the system. The diagnosis accuracy would become better and better with updating. In fault diagnosis, symptoms that are revealed from observations and monitoring data are substantial evidence to diagnose the fault of dynamic systems. However, in practice, the inevitable noise corrupts the symptoms, making it difficult to diagnose the fault accurately in the early stage. Therefore, filtering the noise is the main task for improving the diagnostic accuracy in the incipient fault diagnosis.

The cumulative diagnosis methods can filter the noise and then improve the diagnosis accuracy. However, it remains some challenges. It is difficult to determine the parameters or prior knowledge of the cumulative average based methods. For example, in the MA, the time-window width and the data weights are difficult to set to obtain optimal results. For the Bayesian-based methods, the prior knowledge needs to be as accurate as possible. The more accurate prior knowledge leads to more accurate diagnosis results quickly. Nevertheless, it is not easy to obtain prior knowledge for some systems, such as a system newly put in service.

（2） Amplifying the fault symptom in the feature

In the aspect of enhancing the signal to noise ratio (SNR), the cumulative diagnosis methods can reduce the random noise and increase the magnitude of the useful signal component. For the incipient fault diagnosis, the useful signal component is an unapparent fault feature that cannot indicate the fault reliably. Cumulative diagnosis methods can obtain more accurate diagnosis results by increasing the magnitude of the unapparent fault feature. In a noisy environment, the feature of an incipient fault is unapparent. It is difficult to detect the incipient fault reliably due to the small fault magnitude and noise influence. In this situation, some cumulative diagnosis methods were proposed to enhance the SNR through amplifying the fault symptom, such as the cumulative sum (CUSUM) chart and its variants [52][53].

The CUSUM chart is a sequential and cumulative analysis method for anomaly detection. E. S. Page [54] presented this approach in 1954, a few years after the method of the SPRT. As its name implies, the CUSUM refers to calculate the sum of the variable value sequentially. Unlike the SPRT, the CUSUM does not require using the likelihood function to calculate the cumulative value. Besides, to reduce the detection delay, this

method uses zero function as a holding barrier rather than that threshold calculated according to two decision risk levels. In this situation, the CUSUM uses only one threshold (upper or low threshold), which is preset based on two decision risk levels. Since the zero function is a holding barrier, the CUSUM detects the change in the process in one direction, i.e., positive or negative change. The change in the positive or negative direction is detected when the cumulative sum crosses the preset upper (or low) threshold.

The CUSUM chart provides a simple method to detect the changes in the mean and variance [55][56]. Thus, the CUSUM chart has been employed widely for early fault diagnosis. Using the CUSUM chart, the incipient fault feature is accumulated and amplified. Therefore, the incipient fault can be detected clearly. In the fields of mechanical and electronic systems, CUSUM is employed for fault diagnosis widely. In the mechanical system, CUSUM combined with both symbolic dynamic filtering and intrinsic characteristic-scale decomposition is proposed to diagnose the incipient fault of the bearing [57]. In the Tennessee Eastman process, the CUSUM-based method detects and diagnoses some faults which have not been diagnosed by other algorithms [52][58].

To improve the performance of the CUSUM, researchers present some variants, such as fast initial response CUSUM [59], combined Shewhart-CUSUM [60][61], and multivariate CUSUM [62].

In the CUSUM, the value of the CUSUM is set or reset to zero at the start-up. To obtain a fast initial response feature, Lucas and Crosier proposed to use an initial head start value rather than zero [59]. The initial head start value will get an out-of-control signal fast comparing with the incipient CUSUM value of zero if the change in process is not significant. However, for the process starting out-of-control, the initial head start value has little effect.

The CUSUM detects the small persistent changes in the process optimally. However, for the significant changes, the CUSUM is not as fast as the Shewhart control chart. To improve good control quality, the combined Shewhart-CUSUM was presented [60][61]. The CUSUM is used to quickly detect small persistent changes, whereas the Shewhart control chart is used to detect the significant changes quickly. Thus, the Shewhart-CUSUM obtains quick detection of both small persistent and significant changes.

The CUSUM mentioned above is used for the univariate signal. For complex systems, one sensor cannot catch enough information for accurate fault diagnosis. Nowadays, with the development of the sensors, multiple sensors are installed in one system for obtaining multivariate signals simultaneously. To analysis the obtaining multivariate signals, some researchers extended the univariate CUSUM to the multivariate CUSUM [62]. Comparing with the Hotelling multivariate control-chart, which makes decisions only based on the most recent observed data, the multivariate CUSUM accumulates the information hidden in the historical data and the most recent observed data [63]. More information is obtained from more data. Since the multivariate CUSUM is updated at each time based on historical data and current data, it is more sensitive to the incipient faults than the Hotelling multivariate control chart.

The CUSUM chart based methods consider the current and all the previous data instead of the current data only. Due to considering more information in measuring data, the CUSUM chart based methods are powerful for detecting the incipient fault. In CUSUM chart based methods, before starting a new program, one must establish the acceptable and non-acceptable standard limits for making decisions. In the case of a small amount of data, it is challenging to establish acceptable and non-acceptable standard limits.

（3） Filtering or reducing some frequency components residing in the signal

The cumulative methods also can filter some specific frequencies. In nature and engineering fields, periodicity exists widely. For the purpose of analyzing the signal, some undesired specific frequencies need to be filtered. The cumulative diagnosis methods can implement this task based on the periodicity of the signal. Nowadays, several cumulative diagnosis methods have been developed, such as SMA, response integral quantity method, and TSA.

The principle of this kind of cumulative diagnosis method is to calculate the average or integral or sum over a specified period corresponding to the specific frequency and its harmonics to detect and identify a fault. The SMA aims to calculate the average of the values in the specified period for detection purposes. The length of the SMA is set to the sample number in the specified period. The obtained average will be constant if the signal only includes the specified frequency and its harmonics. Otherwise, the obtained average is not constant. Based on this property, the SMA was used to filter the specified frequency and its harmonics [64][65]. After filtering the specific frequencies, some components can be highlighted, such as fault components. The highlighted fault components are convenient to use for systems monitoring. For instance, to detect chatter, Ma et al. employed the SMA to remove the periodic component relating to the spindle revolution in the measured force signal [64]. After filtering the specific frequencies and random noise, the SNR of the chatter signal is magnified for chatter detection.

Different from the SMA, the response integral quantity method is to calculate the integral of the signal in the designative period [66]. According to the filtering demand, the integration interval is set to nT, where the T is the period of the excitation, and n is a positive integer. The frequency of $1/(nT)$ and the corresponding harmonics are filtered by using the response integral quantity method. This method is used to distinguish period-N bifurcations and chaos [66]. In comparison with the Poincaré section, this method allows identifying the high-order responses of the period-N bifurcations. Because some fixed points in the Poincaré section may be closed, it is difficult to distinguish them. Therefore, the high-order responses of the period-N bifurcations are not identified clearly by using the Poincaré section. Besides, the computing cost of the response integral quantity method is not expensive.

The specific frequencies can be filtered by both the SMA and the response integral quantity method. These two cumulative methods filter the specific frequencies by accumulating the data one by one recursively with the designative time-window width. According to the periodicity of the signal, the cumulative method can also be conducted by accumulating the data one period by one period synchronously. A popular time-

domain signal processing method, TSA, is presented based on this approach [67]. Equation (2-1) shows the TSA. If N is sufficiently large, the TSA signal $\bar{s}$ closely approximates the periodic waveform with the periodicities corresponding to the accumulating period T and its harmonic periods. Therefore, other sources are filtered. Besides, other sources also can be obtained by subtracting the TSA signal $\bar{s}$ from the original signal. Based on the TSA signal $\bar{s}$ or other sources, the fault diagnosis can be conducted.

The SMA and response integral quantity method filter the specific frequency and its harmonics through accumulating the data according to the periodicity. In contrast, the TSA filters the asynchronous components with the reference and extracts the synchronous periodic waveform from the measurement signal. The computational cost of these three methods is cheap. Therefore, these three methods can be used for real-time fault detection and diagnosis. To implement these three methods, we should collect the signal synchronously. In practice, it is difficult to obtain the synchronous signal in some cases. In this situation, these three methods are limited.

（4） Consolidating the degradation information

In addition to the three functions abovementioned, the cumulative diagnosis approach can also continue degradation information accumulation and consolidation. To implement this function, researchers proposed several methods through accumulating the information from current and historical data.

In practice, the evolution of the degradation process of dynamic systems is complex. For instance, according to [4], the bearing may have a function of "self-repair." The phenomenon of "self-repair" expresses that the characteristic of a fault indicator is not monotonic. In this situation, it is difficult to identify the size of the fault. In this case, the cumulative diagnosis approach can accumulate the degradation information and improve the accuracy of the fault identification [68].

In nonlinear dynamic systems, the once sampling per revolution (OSPR) is applied widely for investigating the bifurcation and chaos [17][69][70]. The once sampled per revolution data is constant for the stable process while has fluctuation for the unstable process (i.e., bifurcation and chaos). However, a set of once sampled per revolution data cannot depict the bifurcation character adequately because each set of once sampled per revolution data only depicts partial bifurcation information [16][71]. For the Hopf bifurcation, Zhao et al. [71] showed the difference between several sets of once sampled per revolution data. For a milling process signal, several sets of once sampled per revolution data, which are close to zero, have smaller fluctuations than those far from zero. This phenomenon is the physical nature of the nonlinear models. Kolokolov and Monovskaya [72] reported a similar phenomenon. For the period-2 bifurcation, different sets of once sampled per revolution data express quite different information. For instance, one set of once sampled per revolution data shows the characteristic of a stable process. However, another set of once sampled per revolution data shows the characteristic of period-2 bifurcation. In this situation, it is easy to find the limitation of the once sampled per revolution data for early chatter detection. For the same chatter signal, the different datasets sampled by the once sampling per revolution express different degrees of

fluctuation. Thus, the dataset sampled by the once sampling per revolution fails to depict the degradation character adequately. Moreover, the selection of a dataset sampled by the once sampling per revolution affects the accuracy of the degradation detection. In other words, in a noise background, the degradation is not detectable reliably using one set of once sampled per revolution data. Aimed at this issue, Zhao et al. [71] proposed to consolidate the chatter information through accumulating the information from several sets of once sampled per revolution data.

Besides, the reinforced learning scheme was also proposed for consolidating the degradation information. Generally, the reinforced learning scheme is an evolutionary strategy that generates improved results progressively. The Bayesian belief network (BBN) and genetic algorithm (GA) are representative of the reinforced learning scheme in fault diagnosis. These two methods are widely proposed for fault diagnosis and maintenance [73][74][75][76]. For instance, Adjallah [7] proposed to use the Bayesian belief network for reducing the cost of diagnoses and improving the systems' availability. Nguyen et al. [77] proposed to use the GA to select the optimal features for fault classification. For other fault diagnosis based on the reinforced learning scheme, see the references [73][78][79][80].

We introduced four functions of the cumulative diagnosis approach. The incipient fault can be detected by using the cumulative methods. In general, the more adequate and accurate data allows us to obtain more accurate fault information, and then more accurate diagnosis results can be obtained. Thus, the cumulative methods can diagnose the incipient fault accurately by accumulating the historical signal data.

2.3 State of the art on early-chatter detection in the high-speed milling process

In the modern aerospace and automotive industry, the high-speed milling machine has been widely employed to improve material removal rate and to obtain high-quality machining surface. In high-speed milling processes, one of the most common phenomena to be handled is the chatter, which is a self-excited vibration phenomenon causing instability in the cutting process [18]. The appearance of the chatter reduces productivity, accelerates wear of the cutting tool, leads to low machining efficiency, and increases the machining cost due to loss of quality. The early-chatter should be detected as soon as possible to avoid the damage of the work-piece and save time in the chatter control loop [19].

The chatter can be detected by identifying the chatter features and monitoring the transition between the chatter-free and chatter stages. As a general scheme of fault detection, the procedure of early-chatter detection also consists of three steps: data acquisition, feature extraction, and decision-making. To acquire valid data, we need to determine suitable sensors. The data, incorporating the information of the early-chatter, are sampled. After acquiring data, one extracts the feature for detecting the appearance of a chatter. Using the extracted feature and a preset threshold, one can detect the early-chatter in the step of the decision making. Therefore, we review the early-chatter detection from data acquisition to feature extraction and

decision-making.

2.3.1 Data acquisition

Up to now, several types of sensors are adopted to acquire the data of chatter related signals. Data acquisition is the first step for early-chatter detection. The data acquisition aims to obtain data incorporating the chatter information. The more chatter information captured in the step of the data acquisition allows extracting a better chatter feature. Therefore, various types of sensors were investigated for capturing the chatter information in chatter detection. Commonly used sensors include acceleration sensors, force sensors, displacement sensors, acoustic emission sensors, and so on [18][19][81][82]. The authors in [19][81] have compared and identified the relative strengths and weaknesses of different sensors in chatter detection. Since the chatter is a self-excited vibration phenomenon, it is intuitive to employ the sensors to measure the vibration signal. The vibration signal can reflect the dynamic behavior of a cutting process directly and has been commonly used in chatter detection [83]. Thus, we propose to use the vibration signal to detect chatter in this thesis.

In the step of data acquisition, the study on sampling strategies is the other topic. In chatter detection, the sampling strategies can be categorized into time-domain sampling and angular-domain (or synchronous) sampling [84]. The time-domain sampling is to sample data or signal at an equal-time interval or at constant time increment regardless of the variation of the signal frequency. In fault diagnosis, many researchers from both academic and industrial fields conduct data acquisition by time-domain sampling. To obtain accurate diagnosis results, researchers proposed several spectral analysis methods to transform the waveform signal from the time-domain to the frequency-domain spectrum. In stable signal frequency conditions, the transformed spectrum can reflect the frequency information of the time-domain signal. However, the traditional spectral analysis methods may not work well for the signal with the time-varying frequency since the time-varying frequency leads to the appearance of the smearing phenomenon in the spectrum. For the purpose of addressing this issue, angular-domain sampling was proposed for fault diagnosis of rotating machines.

Angular-domain sampling is also called synchronous sampling. This strategy samples the data or signal at an equal-angle increment. Therefore, the sample number is constant in one revolution. Compared with the time-domain sampling, the number of angular-domain sampling increases with the increase of the rotating speed. Therefore, the smear can be suppressed in the spectrum.

There exist many angular-domain sampling techniques in practice. Referring to Lu et al. [84], these techniques can be classified into three categories: hardware angular-domain sampling, computed angular-domain sampling, and tacholess angular-domain sampling. The hardware angular-domain sampling aims to obtain synchronous data by using hardware equipment. The hardware equipment includes an adjustable analog filter, analog-to-digital converter, and microcontroller unit associated with a sensor. The

microcontroller unit controls the adjustable analog filter and the analog-to-digital converter to sample the data synchronously in variable-speed conditions. The computed angular-domain sampling aims to transform the time-domain sampling data to the angular-domain sampling one by using the numerical interpolation-based signal resampling. In the process of transformation, the tachometer signal is used as the reference signal. In the tacholess angular-domain sampling, there is no reference signal. Therefore, the rotating speed needs to be estimated from the available signals. After estimating the rotating speed, the computed angular-domain sampling can be conducted to obtain the synchronous data. From which, we can regard the tacholess angular-domain sampling as a special computed angular-domain sampling independent of using a tachometer.

In data acquisition, the different types of synchronous data can be attained according to different sampling strategies for various applications. In practice, the synchronous data has been used to various topics such as order tracking [84], rotating speed estimation [85], preventive diagnosis [86], milling stability interrogation [87], and early-chatter detection [71]. To apply the synchronous data to various applications, researchers proposed some sampling strategies, such as general angular-domain sampling, once sampling per revolution, and subharmonic sampling.

The general angular-domain sampling aims to collect the synchronous data at the equal-angle interval or constant angle increment. The angle interval is commonly less than 2π since the sampling frequency needs to be higher than twice the maximum frequency of the signal for signal reconstruction according to the Nyquist-Shannon sampling theorem. Otherwise, the signal cannot be reconstructed. The angular-domain sampling with small-angle intervals is used for the spectral analysis and order tracking widely.

The once sampling per revolution collects synchronous data through angular-domain sampling at a fixed-angle interval of 2π [88]. This sampling strategy is derived from the Poincaré map technique, which has been widely applied to instability research in nonlinear systems [89][90]. As the name suggests, the once sampling per revolution means to acquire one sample in one revolution synchronously. The once sampling per revolution technique reduces the amount of the signal significantly compared to the traditional sampling approaches. Specifically, one sample is acquired during one spindle revolution synchronously. However, for the order tracking, tens of kilohertz sampling rates might be required to avoid aliasing, which might occur in the signal translation process [88]. Generally, less data usually needs less computational cost for signal processing and data transmission. So, this technique could reduce the computational cost of signal processing and also can be used for remote data transmission for remote chatter detection, fault diagnosis, and system control. Based on the OSPR technique, Schmitz et al.[88][91] presented an online chatter detection method. The distribution of the once sampled per revolution data is different at the chatter and chatter-free stages. Honeycutt and Schmitz [92] used the once sampled per revolution data for automatic stability identification.

However, a set of once sampled per revolution data cannot adequately depict the chatter character because

each set of once sampled per revolution data only depicts partial chatter information [16]. For the quasi-periodic chatter associated with the Hopf bifurcation, each set of once sampled per revolution data has fluctuation, which can be used to depict the appearance of the chatter. Conversely, for the chatter associated with the period-2 bifurcation, there is a set of once sampled per revolution data without the fluctuation [16]. The chatter-free signal and the period-2 chatter signal cannot be distinguished based on this set of once sampled per revolution data. Thus, the early-chatter detection methods based on the OSPR technique need further development for practical applications.

In time-varying rotating speed conditions, the time-domain sampling can lead to the smearing phenomenon in the spectral analysis. Nonetheless, the general angular-domain sampling overcomes this disadvantage of the time-domain sampling for the spectral analysis. Besides, the OSPR can filter the rotating frequency and its harmonics and detect some kinds of chatter. Nevertheless, once sampling per revolution may fail to detect the period-2 chatter. The OSPR needs further development for obtaining reliable chatter detection results.

2.3.2 Feature extraction

Among the three steps of chatter detection, feature extraction is the most important one. Therefore, significant research works have been proposed to extract reliable chatter features based on signal processing methods and mathematical models of chatter physics. Table 2.1 shows the basic performances required from chatter detection methods in the literature. None can detect chatter at an early-stage with a theoretical basis and dependence on machine parameters.

Table 2. 1 Comparison results of some chatter detection methods

		Researches	Ability to detect chatter in the early stages	Dependency on machining parameters	Theoretical basis
Methods based on signal processing methods	Time domain	Ref. [93]	**Yes**	**Yes**	No
		Ref. [94]	**Yes**	No	**Yes**
	Frequency domain	Ref. [95]	**Yes**	No	**Yes**
		Ref. [96]	No	No	**Yes**
		Ref. [99]	No	**Yes**	**Yes**
	Time-frequency domain	Ref. [15]	**Yes**	No	**Yes**
		Ref. [97]	No	No	**Yes**
		Ref. [98]	No	No	No
		Ref. [100]	No	No	**Yes**
Methods based on mathematical models of chatter physics	Signal model	Ref. [101]	No	No	**Yes**
		Ref. [102]	**Yes**	No	**Yes**
	Dynamic state model	Ref. [85]	No	**Yes**	**Yes**
		Ref. [103]	No	**Yes**	**Yes**
		Ref. [88]	No	**Yes**	**Yes**
		Ref. [91]	No	**Yes**	**Yes**
		Ref. [92]	No	**Yes**	**Yes**
	Probabilistic model	Ref. [104]	No	**Yes**	**Yes**
		Ref. [105]	No	**Yes**	**Yes**

The chatter signal during machining has complex nonlinear and non-stationary characteristics. So, the

authors employed nonlinear and non-stationary signal processing methods in the time domain [93][94], frequency domain [95][96], and time-frequency domain [15][97][98] to detect the early-chatter.

In the time domain, the chatter feature can be extracted by calculating the statistical characteristics of the collected signal, such as mean, variance, and standard deviation [93]. The self-organizing map (SOM) neural network is proposed for multi-feature fusion. The developed chatter indicator can detect early-chatter by comparing it with a preset threshold. Nevertheless, the application of this early-chatter detection method is limited due to no theoretical basis. Jia et al. [94] presented a synthetic criterion (SC) by integrating standard deviation and one-step auto-correlation function for early-chatter detection. The experimental results verified that SC could detect early-chatter. However, the variation of machining parameters deteriorates the chatter detection performance of the SC.

In the frequency domain, based on the energy aggregation property of the chatter, Hynynen et al. [95] proposed to use the coherence function of the acceleration signal and the audio signal to detect the early-chatter. The experimental results showed that the proposed method is sensitive to early-chatter. Nevertheless, this method is not independent of machining parameters, so the optimal threshold must be updated with the machining parameters variation. Addressing this problem, Aslan et al. [99] and Cao et al. [93] have explored the independent chatter detection methods and suggested using the comb filter. The tooth passing frequency, spindle rotation frequency, and corresponding harmonics are filtered from the chatter signal due to employing this filter. Thus, the chatter feature based on the filtered chatter signal is independent of machining parameters. However, it is a challenge to determine the number of spaced notches of the comb filter [64]. This disadvantage limits the application of comb filter-based methods in early-chatter detection.

In the time-frequency domain, wavelet transform (WT), empirical mode decomposition (EMD), and variational mode decomposition (VMD) are standard signal processing methods to detect an early-chatter. These advanced signal processing methods decompose the chatter signal into some components/modes corresponding to different frequency bands. The decomposition components/modes, including rich chatter information, can serve to highlight and extract the chatter features. However, WT-based and VMD-based chatter detection methods need to determine optimal decomposition parameters in advance. In practice, an unsuitable choice of decomposition parameters affects the detection effectiveness. Up to now, there is no selection criterion of the optimal decomposition parameters. Although EMD-based chatter detection methods do not need to determine optimal decomposition parameters in advance, their applications are limited due to no strengthened theoretical basis [100].

The abovementioned chatter detection methods rely on signal processing and do not consider the mathematical model of chatter physics. Some chatter detection methods using the mathematical model of chatter physics were also introduced for real-time implementation. The mathematical model of chatter physics can be categorized into signal model [101][102], dynamic state model [85][103], bifurcation and

chaos model [17][88][91][92], and probabilistic model [104][105].

Some researchers adopted the nonlinear process to model the milling process time-varying dynamics due to the milling process signal nonlinear characteristics. The autoregressive moving average (ARMA) model and exponential Autoregressive (ExpAr) model are typical. Van Dijk et al. [101] proposed a parametric modeling approach for detecting the chatter in the milling process. The parametric modeling approach utilizes the ARMA model of the signal. The proposed model-based method can detect chatter faster in a high-speed milling process. However, it gives no solution to determine the order of the ARMA model. The estimation results may be unsatisfactory if the chosen order of the ARMA model is not suitable. Similarly, the inappropriate parameters of the ExpAr model can also lead to disappointing results [102].

Dynamic state models, incorporating the milling process dynamic properties, were also employed to detect chatter. For instance, Kakinuma et al. [85] designed a disturbance observer based on the spindle current and speed signals to detect chatter. The chatter and forced vibrations are separated by using digital filters. They verified the effectiveness of the disturbance observer using several milling tests. However, there is no principle to set the frequency bands of the designed digital filters. Caliskan et al. [103] proposed a Kalman filter to detect and extract chatter components. The Kalman filter is assigned to estimate the forced vibration component. Then the chatter vibration component can be obtained by subtracting the forced vibration component from the measured signal. However, the measured signal is nonlinear and non-stationary. The chatter vibration component obtained using the Kalman filter may not adequately depict the chatter character since the Kalman filter's ability is limited for processing nonlinear and non-stationary signals.

The chatter is an unstable phenomenon also related to bifurcation or chaos. Thus, some researchers addressed the chatter detection problem based on the bifurcation and chaos model [106][89][107]. Schmitz et al. [88] used the OSPR technique to detect chatter. Based on the property of the milling signal with a constant revolution speed, one can find that the once sampled per revolution data is constant at the chatter-free stage. However, the once sampled per revolution data has fluctuations at the chatter stage. In theory, due to the different distributions of the once sampled per revolution data at chatter-free and chatter stages, the chatter can be detected easily [88][91][92]. The once sampled per revolution data is independent of the machining parameters and can be used to detect chatter.

In practice, the measured signal includes noise due to some uncontrollable influences, such as the measurement instrument and in-homogeneous material of the work-piece [108]. Face to this situation, the boundary between chatter-free and early-chatter is fuzzy. It is difficult to extract the chatter feature for detecting early-chatter. We considered this problem in this work, where we propose a cumulative early chatter detection method based on the OSPR technique and sequential probability ratio test. The early chatter was detected by accumulating the unobvious chatter feature. Besides, this method requires a small amount of sampled data by using OSPR. Less data requires less cost for data transmission. Therefore, the proposed early-chatter detection method is suitable for implementing a wireless sensor for remote

monitoring and early-chatter detection.

Many random factors affect the chatter appearance in the milling process, such as machine degradation, uncertainties of machining parameters, and in-homogeneous material [109]. Therefore, the probabilistic process model is suitable for chatter detection [104][105][71]. The probability distribution of the cutting force under the stable cutting conditions is different from that obtained under the chatter conditions in the milling process. Based on this phenomenon, Du et al. [104] proposed to detect chatter by using the variance ratio of the probability distribution of the cutting force. Although this method can detect severe chatter, the early-chatter could not be detected due to the fuzzy boundary between chatter-free and early-chatter. There is a transient process between the stable cutting process and the chatter process during the chatter appearance. The wavelet transform's modulus maxima (WTMM) approach effectively detects the transient phenomena in image enhancement [110]. Therefore, Wang and Liang [105] proposed to detect chatter based on the probability distribution of WTMM. The authors demonstrated that the probability distribution of WTMM belongs to the Weibull distribution. The chatter feature is extracted based on the probability distribution of WTMM. Experiment results valid this method can detect chatter in the constant and variable speed regimes. Nevertheless, the early-chatter may not be detected due to the inapparent chatter feature.

The chatter feature extraction methods are reviewed. Real-time implementation is one of the important requirements of early-chatter detection since the time from chatter onset to mature chatter is short. Besides, the generalized early-chatter detection methods are worth exploring for implementing chatter control [111].

2.3.3 Decision making

In chatter detection research works, the strategies for data acquisition and feature extraction were investigated widely. However, in the aspect of the decision-making, the studies of thresholds are rarely reported.

Most of the chatter detection works focus on judging the appearance of the chatter based on the traditional threshold constant value [19][112]. The occurrence of chatter is recognized when the chatter index exceeds the preset threshold value [112].

However, in practice, the chatter index may not exceed the threshold value after the appearance of the early-chatter due to its random property. Also, the threshold value may be exceeded when the machining process is stable. These two types of decision risk cannot be avoided in decision making based on the traditional threshold value. In other words, the threshold value fails to detect the early-chatter reliably. To detect early-chatter reliably, we should propose another type of threshold.

2.3.4 Summary

Early-chatter detection is a critical approach to ensure machining quality and efficiency in high-speed milling processes. In this section, we review the early-chatter detection from the data acquisition, feature extraction to the decision making. Two performance requirements of early-chatter detection are to detect chatter

quickly and reliably. In other words, real-time early-chatter detection, which is independent of machining parameters, is required in the high-speed milling process.

Some research works must be conducted to obtain quick and reliable early-chatter detection results. In the step of data acquisition, one should acquire data that are independent of machining parameters and includes sufficient chatter information. In the step of the feature extraction, the reliable chatter feature should be extracted in real-time. In the decision-making step, another type of threshold must be proposed to overcome the limitation of the traditional threshold value. To implementing early-chatter detection quickly and reliably, one should propose more efficient real-time early-chatter detection methods.

2.4 State of the art on period-*N* bifurcations identification

In general, early-chatter refers to the phenomenon of bifurcation. In the machining process, the common bifurcation is the period-*N* bifurcations and secondary Hopf bifurcation [13]. Referring to Honeycutt and Schmitz [13], the machining quality under period-2 bifurcation conditions can be better than that under secondary Hopf bifurcation conditions. Therefore, to achieve higher productivity while still satisfying the machining quality requirements, the operator may choose the machining parameters causing the period-*N* bifurcations rather than ones causing the secondary Hopf bifurcation.

In practice, the changes in the natural frequency of the machining system or the machining parameters can switch the machining behavior from the period-2 bifurcation to the secondary Hopf bifurcation. So it is necessary to identify the period-2 bifurcation to improve the material removal rates while still satisfying the machining quality requirements.

In addition, all dynamic systems operating at a given type of bifurcation exhibit similar characteristics. However, systems operating at a different type of bifurcation express a different character. Generally, the bifurcation presence refers to a system fault or function fault in a mechanical and electrical system. Therefore, by identifying the bifurcation type and size, one may determine the fault type and size or predict the system performance [113]. Therefore, period-*N* bifurcations identification has great significance in practice.

Up to now, some research works have been reported in the identification of period-*N* bifurcations or measurement of sub-harmonics. The sub-harmonics are defined as non-integer harmonics with the frequency below the fundamental frequency [114] and are relative to the period-*N* bifurcations. Among these research works, commonly used methods are Fourier transform (FT) [115], MUSIC method [114], wavelet-packet transform (WPT) [116], Poincaré map [13], response integral quantity method [66], once-per-revolution sampling [117] and sub-harmonic sampling [87]. Therefore, we review these methods abovementioned, respectively.

The FT is a well-known linear signal processing method. It can transform the signal from the time-domain into the frequency-domain. In the frequency domain, the signal frequencies are illustrated clearly. In power

systems, Barros et al. [115] measured sub-harmonics by using discrete Fourier transform (DFT). Similarly, the IEC Standard 61000-4-7 measures harmonics and inter-harmonics using DFT in power supply systems. The time-window width determines the resolution of DFT. In the IEC Standard 61000-4-7, the chosen time-window width and the fundamental frequency of the response are ten fundamental periods and 50Hz, respectively. Therefore, the resolution of DFT is 5Hz. In some situations, the resolution of 5Hz is too rough and not satisfactory. Increasing the time-window width can improve the FT resolution. However, the extended time-window width leads to a long delay in measurement. Moreover, non-stationary phenomena may occur in a wide time-window and then smear the FT performance [114].

Aimed at the low resolution of FT, Leonowicz proposed analyzing the sub-harmonics using the root-MUSIC method [114]. The root-MISIC method has a higher resolution and better accuracy than the FT in most cases. Nonetheless, the computational cost of the root-MISIC method is more expensive than that of FT. The expensive computational cost is not appropriate for the implementation of the real-time identification of period-N bifurcations.

WPT is proposed to measure the non-stationary and non-synchronous sub-harmonic. WPT is a nonlinear and non-stationary signal processing method. Diego and Barros [116] compared the performance of DFT and WPT in measuring sub-harmonics under three possible measurement conditions. They used three examples to represent three possible measurement conditions: synchronous stationary sub-harmonic, non-synchronous stationary sub-harmonic, non-synchronous, and non-stationary sub-harmonic. The comparing results of those three examples were summarized. The DFT has excellent performance for stationary and synchronous sub-harmonic, whereas the WPT shows excellent performance at the condition of the non-stationary and non-synchronous sub-harmonic. In practice, the decomposition parameters of WPT should be determined in advance. An unsuitable choice of decomposition parameters affects measurement effectiveness. Up to now, there is no selection criterion of the optimal decomposition parameters.

Among identification methods of period-N bifurcations or sub-harmonics, the Poincaré map is a commonly used qualitative approach. The Poincaré map indicates the property of the nonlinear system in a certain lower-dimension [13]. In the Poincaré map, N fixed points correspond to the period-N bifurcations. According to the number of fixed points, the period-N bifurcations can be identified easily. However, the Poincaré map may fail to recognize the period-N bifurcations when N is significant because the sampling points shown in the Poincaré map may be very close to each other. Aimed at this problem, Jeng et al. proposed a response integral quantity method [66]. The response integral quantity method considers the periodicity of period-N bifurcations response and integrates the period-N bifurcations response in N times periods of excitation. For the period-N bifurcations response, the integration value is a fixed constant in the N times periods of excitation. Otherwise, the integration value is not equal to a fixed constant. According to this property of period-N bifurcations, the type of period-N bifurcations was identified. However, the identification method in [66] does not consider the noisy situation. Under the influence of the noise, the

integration value is not a fixed constant.

Based on the Poincaré mapping technique, a typical synchronous sampling technique, i.e., once sampling per revolution (OSPR), was derived for chatter detection [118][88]. The principle of OSPR is to collect one sample in one spindle revolution synchronously. The OSPR is a concept in the mechanical field. One sample is collected in one period of excitation or one typical minimum period of multiple excitations for other fields. For instance, in power systems, since the frequency of the alternating current is 50Hz, the period is 0.02s. For convenience, we introduce this typical synchronous sampling using the term in the mechanical field. For other fields, the typical synchronous sampling can be termed according to the applications. Kolokolov and Monovskaya [117] built an amplitude space based on synchronously sampled points in power systems. The maximum distance between the currents of synchronously sampled points represents the X-axis. The voltage represents the Y-axis. In the built amplitude space, different period-*N* bifurcations motions locate different areas. Therefore, the period-*N* bifurcations can be identified based on the amplitude space.

Based on the idea of OSPR, Faulkner, and Mestre [119] presented the sub-harmonic sampling technique. Unlike OSPR, the principle of sub-harmonic sampling is to collect one sample in several revolutions synchronously. To identify the type of period-*N* bifurcations, Honeycutt, and Schmitz [87] proposed to employ the sub-harmonic sampling technique. The period-*N* bifurcations response has *N* times as many as a fundamental period. Thus, the sub-harmonic sampled data corresponding to the period-*N* bifurcations is constant. Based on this property, the type of period-*N* bifurcations can be identified. However, the sub-harmonic sampling may also fail to identify the type of period-*N* bifurcations because each set of sub-harmonic sampling data only depicts partial information of period-*N* bifurcations. For instance, the period-2 and period-4 bifurcation may not be identified by using sub-harmonic sampling. This limitation of sub-harmonic sampling is similar to the disadvantage of OSPR, as will be shown later in Chapter 3. Since a set of OSPR data or sub-harmonic sampled data only contains partial information of period-*N* bifurcations, then the type identification methods of period-*N* bifurcations based on sub-harmonic sampling do not identify period-*N* bifurcations reliably.

The exploration of the data resource may enhance the reliability in the type identification of period-*N* bifurcations. For the purpose of acquiring the information of the high order harmonics, the sampling frequency is set much higher than the revolution frequency. Thus, there are some samples in one fundamental period. In this situation, synchronous data can be obtained. Generally, the more data is considered, the more reliable the result will be obtained. Therefore, the period-*N* bifurcations identification based on more synchronous data may be a possible way.

The simple moving average (SMA) is a synchronous averaging technique for filtering the noise and can identify period-*N* bifurcations by accumulating the synchronous data. In practice, the signal includes noise. From the above analysis, one can see that it is not easy to identify the incipient period-*N* bifurcations using

one set of sub-harmonic sampled data. For improving the reliability in the type identification of period-*N* bifurcations, one promising way is to filter the noise based on the synchronous data. SMA is one of the most commonly used time-domain filters. The SMA can filter the noise and remove the specific frequency and harmonics completely by accumulating the history data. Due to the property of filtering specific frequency and its harmonics, the SMA can be considered as a comb filter [20]. In power systems, the SMA has been used for harmonic identification and elimination [120][121]. Since period-*N* bifurcations response has corresponding period components, the SMA seems to be able to identify the type of period-*N* bifurcations suitably. In this thesis, we will propose to identify the type of period-*N* bifurcations based on the SMA.

Up to now, few works on size identification of period-*N* bifurcations are reported [116]. In practice, it is necessary to identify period-*N* bifurcations size for evaluating the level of period-*N* bifurcations. In order to fill this gap, this thesis will present a new method to identify the size of period-*N* bifurcations.

2.5 Conclusions

State of the art on cumulative diagnosis, early-chatter detection, and period-*N* bifurcations identification are introduced, respectively. The cumulative diagnosis approach has four advantages: (1) reducing the random noise residing in the signal or incipient fault feature; (2) amplifying the fault feature, (3) filtering or reducing some frequency components residing in the signal; (4) consolidating degradation information. The early-chatter detection and period-*N* bifurcations identification, as the applications of the cumulative diagnosis approach, are also reviewed. In the early-chatter detection and period-*N* bifurcations identification, quick and reliable detection and identification results are necessary. The cumulative diagnosis methods are potential tools to address this issue. Therefore, we will propose three early-chatter detection methods and a period-*N* bifurcations identification method based on the cumulative diagnosis approach. The proposed methods will be introduced in chapters 3 and 4, respectively, while considering chatter faults in milling as an application.

Chapter 3 Early-chatter detection methods based on cumulative diagnosis approach

3.1 Introduction

In early-chatter detection, reliable and quick detection of the early-chatter is a challenge. As mentioned in chapter 2, the early-chatter feature is incipient and buried into the noise. Thus, it is challenging to detect early-chatter reliably. Besides, the time from chatter onset to mature chatter is short (less than 0.1 s) in the high-speed milling process. It is also challenging to detect early-chatter real-time in such a short period. The cumulative diagnosis approach has the potential to address these two issues. Because the cumulative diagnosis approach has four advantages: (1) reducing the random noise residing in the signal or incipient fault feature; (2) amplifying the fault feature; (3) filtering, or reducing some frequency components residing in the signal; (4) consolidating degradation information. Therefore, this chapter proposes three early-chatter detection methods based on the cumulative diagnosis approach.

It is similar to the fault detection methods based on signal processing. The proposed early-chatter detection methods also include three main steps: data acquisition, feature extraction, and decision-making. Therefore, this chapter first introduces the basic elements for early-chatter detection in terms of data acquisition, feature extraction, and decision-making. Then, we propose three early-chatter detection methods, respectively, and lastly, draw some concluding remarks.

3.2 Basic elements of early-chatter detection

3.2.1 Elements of data acquisition

Data acquisition is the first and essential step for early-chatter detection. Data acquisition is a process of collecting and storing data or information. The collected and stored data or information is used for feature extraction. The data acquisition can be implemented using the sampling instruments in two sampling domains: time-domain and angular-domain. The corresponding acquired data is called time-domain sampled data, or angular-domain sampled data (dynamics' synchronous data).

The time-domain sampled data is collected at a specific time interval. In practice, setting a specific time interval is easy. Thus, this type of data acquisition is the most commonly used. However, for some dynamic systems, such as rotating machines, reciprocating machines, power supply systems, telecommunications,

and electronic or electromagnetic systems, under variable operating frequency conditions, the time-domain sampled data yields non-stationary signals. Due to the effect of the operating frequency variation, smearing phenomena can appear when the time-domain sampled data are transformed into the frequency domain by using the fast Fourier transform (FFT). The appearance of the smearing phenomenon affects the reflected information in the frequency domain. Aimed at this problem, researchers proposed to use angular-domain sampled data. The angular-domain sampled data are collected at an equal-angle increment. This sampling method enables avoiding smearing phenomena appearing in the frequency domain. That is why the angular-domain sampled data is preferred over the time-domain sampled data in some dynamic systems applications, such as rotating machines, reciprocating machines, power supply systems, telecommunications, electronic or electromagnetic systems.

Since three proposed early-chatter detection methods are based on angular-domain sampled data, we introduce the angular-domain sampling process in chapter 3.

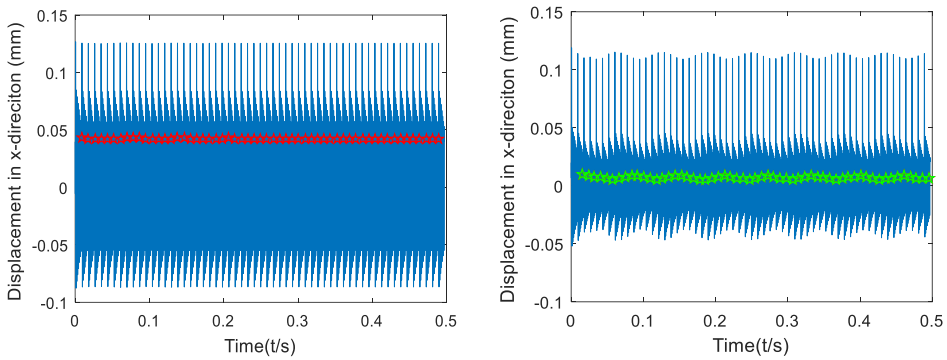
The idea of using angular-domain sampling in the mechanical field was stemmed from the fields of electronics and telecommunications, where several research works have been dedicated to this technical issue starting from the 1960s [122][123][124]. The angular-domain sampled data can be acquired using an encoder or a tachometer installed on the shaft of rotating machines. According to the acquired sample number in one revolution, one can classify the angular-domain sampling into two categories: once sampling per revolution (OSPR) and multiple sampling per revolution (MSPR).

3.2.1.1 Once sampling per revolution

The OSPR refers to collect one sample in one revolution synchronously [89]. In other words, the OSPR data are collected at the equal-angle 2π increment in the angular-domain. This approach is derived from the Poincaré sectioning technique, which has been widely applied to instability research in nonlinear systems [89][90]. The once sampling per revolution technique reduces the amount of the samples of the signal significantly compared to the traditional sampling approaches. For instance, for applying the FFT method, tens of kilohertz sampling rates should be required to avoid aliasing, which might occur in the signal translation process [88]. Generally, less data usually allows a lower computational cost for signal processing and data transmission. So this technique could reduce the computational cost of the signal processing and can be used to remote data transmission for remote chatter detection, fault diagnosis, and system control.

Figure 3.1 illustrates the OSPR data at the chatter-free and early-chatter stages. We can find that the once sampled per revolution data is constant at the chatter-free stage. However, at the early-chatter stage, the once sampled per revolution data has a constant period fluctuation. The early-chatter can be detected based on the different dispersion of the once sampled per revolution data at chatter-free and early-chatter stages. Based on the OSPR technique, Schmitz et al.[88][91] presented an online chatter detection method according to the different dispersion of the OSPR data at the chatter-free and early-chatter stages. Honeycutt and Schmitz [92] used OSPR data for automatic stability identification. The OSPR has achieved decent successes for

chatter detection and stability identification. Therefore, we propose to use the OSPR for early-chatter detection in the first proposed method.



(a) (b)

Figure 3. 1 OSPR - Once sampling per revolution data, (a) chatter-free signal, (b) early-chatter signal.

3.2.1.2 Multiple sampling per revolution

As the name suggests, the MSPR refers to collect multiple samples in one revolution synchronously. In general angular-domain sampling, the acquired MSPR data is seen as a series of data. Based on this series of data, the time-synchronous average and order tracking can be implemented for fault diagnosis [125][126][127][128]. Different from the general angular-domain sampling, we present a new MSPR. The latter is devised to acquire angular-domain sampled data by implementing OSPR at different rotation angles repeatedly and then obtain multiple sampled data sets. This MSPR is a new insight into the angular-domain sampling and can be seen as an extension of the OSPR technique. To distinguish the presented MSPR and the general MSPR, hereafter, we note the one presented as MSPR, whereas the general one is called traditional angular-domain sampling.

Assuming there is a digital signal in the phase domain

$$X = \{x(\varphi_0), x(\varphi_0 + \frac{2\pi}{M}), x(\varphi_0 + 2\frac{2\pi}{M}), \ldots, x(\varphi_0 + (M-1)k\frac{2\pi}{M})\}, (M, k \in N^*), \tag{3-1}$$

where φ_0 represents the initial sampling phase, M is the sample number in one revolution, $\frac{2\pi}{M}$ represents sampling interval, k represents the number of revolutions.

According to the Nyquist–Shannon sampling theorem, the sampling frequency must be greater than twice the highest frequency of the signal. The sampling frequency is generally far higher than the rotation frequency to collect the information, including some high-order harmonics. So, it is possible to collect more samples in one revolution synchronously. The MSPR can be described as follows:

(1) At the initial sampling phase φ_0, use the once sampling per revolution to collect the signal X, which yields a set $\{x(\varphi_0), x(\varphi_0 + 2\pi), x(\varphi_0 + 4\pi), \ldots, x(\varphi_0 + 2k\pi)\}$ of once sampled per revolution data.

(2) At the sampling phases $\varphi_0 + m\frac{2\pi}{M}$ ($m = 1,2,3,4,\cdots,M-1$), collect the signal X by using the once sampling per revolution repeatedly, then M-1 sets of once sampled per revolution data are collected.

After the M implementations of the once sampling per revolution technique, we obtain M sets of once sampled per revolution data. Due to multiple sampling in one revolution, this technique is called MSPR. However, for the traditional angular-domain sampling used for order tracking, the angular-domain sampled signal is the digital signal X. Thus, one can say that the MSPR brings an insight into the angular-domain sampling technique. The MSPR data includes M sets of once sampled per revolution data. In practice, the number M of samples should be set according to the applications. A large sample number allows for obtaining more information and leads to a high sampling cost in terms of energy consumption and noise collection.

The proposed MSPR overcomes the shortcoming of the once sampling per revolution in the aspect of depicting the chatter character. For a displacement signal of the cutting tool, one can obtain MSPR data by using the MSPR. Figure 3.2 illustrates the MSPR data (M=9). Due to the nine samples collected in one revolution, it can be found that the MSPR data includes nine sets of once sampled per revolution data. Compared with the once sampled per revolution data shown in Figure 3.1(b), the MSPR data has more information pertaining to the chatter characteristics. Moreover, in MSPR data, the fluctuations in each data set sampled by the OSPR are different. For instance, several sets of once sampled per revolution data, which are close to zero, have smaller fluctuations than those far from zero. This phenomenon is the physical nature of the nonlinear models [16]. A similar phenomenon was reported in [72]. For the period-2 bifurcation, the different sets of once sampled per revolution data carry quite different information. For instance, one set of once sampled per revolution data shows the characteristic of period-2 bifurcation. However, another set of once sampled per revolution data shows the characteristic of the stable process.

From Figure 3.2, one can find that each set of once sampled per revolution data depicts partial information of the chatter's character. It is easy to find the limitation of the once sampled per revolution data for early-chatter detection in this situation. For the same chatter signal, the different datasets sampled by the once sampling per revolution express different degrees of fluctuation. Thus, the dataset sampled by the once sampling per revolution fails to depict the chatter character adequately. Moreover, the selection of a dataset sampled by the OSPR affects the accuracy of the early-chatter detection. In other words, in noise background, the early-chatter is not detectable reliably using one set of once sampled per revolution data. The MSPR can overcome the shortcoming of OSPR. Therefore, we proposed to use the MSPR for early-chatter detection in the second and third proposed methods.

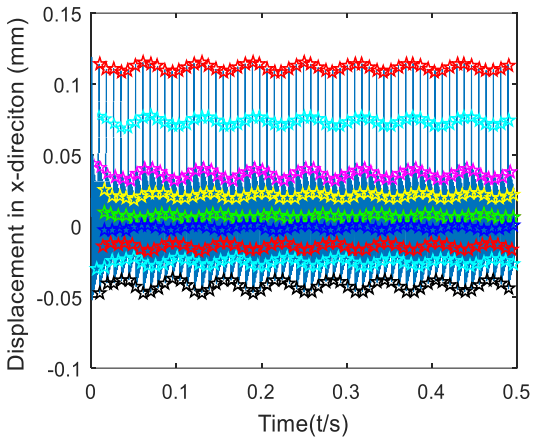


Figure 3. 2 MSPR - Multiple sampling per revolution data (M=9).

3.2.2 Feature extraction

The feature extraction is an essential step for early-chatter detection. An appropriate feature or indicator enables depicting the operating state of the high-speed milling processes. For early-chatter detection, we implemented three early-chatter detection methods using two extracted features based on the probabilistic model. This subsection introduces the two feature extraction methods: 1) the maximum entropy (MaxEnt) principle; 2) the second-order reliability method (SORM).

3.2.2.1 Maximum entropy principle

In 1957, Jaynes introduced the maximum entropy (MaxEnt) principle based on the information entropy introduced by Shannon [129]. The MaxEnt principle indicates that the estimated probability density function (PDF) with MaxEnt is the optimized choice to evaluate a dynamic system behavior-based state in case of insufficient information. Thus, the probability distribution is optimized by maximizing the information entropy. The MaxEnt principle estimates the probability distribution without requiring any specific assumptions. Therefore, the estimated PDF is the maximally unbiased estimation satisfying the available information.

Considering a stochastic variable x, which produces a set of data $x=\{x_1, x_2, \dots, x_i, \dots, x_n\}$, the PDF of x is defined as $f(x)$. Equation (3-2) gives the informational entropy quantifying the uncertainty:

$$H(f(x)) = -\int_{-\infty}^{+\infty} f(x) * \log(f(x))dx \,. \tag{3-2}$$

Equation (3-3) expresses the constraints related to the moments of the stochastic variable x:

$$E(g_j(x)) = -\int_{-\infty}^{+\infty} g_j(x) * f(x)dx = \alpha_j, j = 0,1,2,\dots,m \,, \tag{3-3}$$

where $g_j(x)$ is a specific function relating to an informational component of x, i.e., x^j, and $E(g_j(x))$ represents the statistical expectation of $g_j(x)$. If j=0 then the zeroth moment $E(g_0(x))$ is the total probability, if j=1 then $g_1(x)$=x, $E(g_1(x))$ represents the first-order moment and so forth.

The MaxEnt principle aims to find $f(x)$ without additional assumptions by maximizing the information entropy under the given constraints function of $g_i(x)$ in Equation 3-3. The MaxEnt optimization problem can be solved using the Lagrange multipliers (see [129], [130] for detail).

The MaxEnt principle, with the first four moments, is a mature method and has been widely used to estimate the PDF [131][132][133][134]. In the context of the moment methods, moments incorporate different information about system behavior. However, higher-order moments may lead to the occurrence of instability in estimating the PDF as well. To deal with this issue, Rajan et al. proposed an improved MaxEnt method [135], which can estimate the PDF through accommodating higher-order moments. However, the computational cost is expensive when the MaxEnt principle accommodates higher-order moments. To improve the performance of the MaxEnt principle, some researchers introduced the concept of the fractional moments to estimate the PDF [136][137]. Compared to the MaxEnt principle with integer moments, the MaxEnt principle with fractional moments have lower sampling variability and can optimize the set of moments. Nonetheless, there are several issues in the MaxEnt principle with fractional moments. First, its use is limited in some engineering applications since the fractional moments are not valid for random variables that can take negative values. The second issue concerns the optimization problem. Since the objective function is non-continuous and non-convex, it is difficult to solve the optimization problem [138]. Therefore, we propose to use the MaxEnt feature with the first four moments to estimate the PDF in this work.

3.2.2.2 Second-order reliability method based on MaxEnt feature

In the milling process, many random factors affect the chatter appearance, such as degradation of the machine, uncertainties of machining parameters, and in-homogeneous material [109]. Therefore, the probabilistic process model is suitable for chatter prediction and detection [139].

In industrial engineering, reliability analysis is necessary to ensure the safety and serviceability of systems or parts. Referring to [139] [109], the chatter issue is related to the physic reliability of a work-piece. So, it appears relevant to use the physic reliability-based approach for preventing chatter related failure in a milling process. Such a failure can appear after chatter symptoms before a tool breaks or a product lose its quality during machining. Hence, we consider the physic reliability definition of a work-piece as a probability that the strength of such an item is higher than the load corresponding, in our case, to stresses imposed on it during operation. The illustration in [140][139] describes perfectly the physic reliability of a work-piece in a milling process, where tools, machine components, and work-piece quality could degrade when the strength of a component drops below the stresses or loads. The decrease of the strength may correspond to the tool wearing or the loss of mechanical properties of the work-piece material, or the degradation of the machine component. In reliability analysis, it is a challenge to focus on the design aspect (or the most probable failure point). By applying the second-order reliability method (SORM) based on the maximum entropy (MaxEnt) feature, during machining, one can avoid focusing on design consideration.

Also, the SORM based on the MaxEnt feature can estimate quickly the failure hazard function (FHF), which is used as a chatter indicator. We will introduce FHF in the equation (3-7).

In this section, a reliability analysis method, i.e., SORM based on the MaxEnt feature, is introduced for developing a MaxEnt feature-based reliability model method (MFRM) of real-time detection of early-chatter.

(1) MaxEnt reliability model

The proposed reliability model is the hardware reliability model, according to Rausand [140]. For a structure or system to be reliable, it must withstand the various loads (stresses), be they stochastic or deterministic. Let us consider, for instance, two variables: resistance (noted as *R*) of the structure and load (noted as *S*) on the structure. Here, *R* and *S* are stochastic variables. To satisfy the safety requirement, the resistance *R* should exceed the load *S*, which can express mathematically as

$$Z = R - S > 0, \tag{3-4}$$

where *Z* is the performance feature function.

In practice, the load may increase due to the variation of the operating conditions, or the resistance may decrease due to structural performance degradation. Taking into account all these factors of physical influence, the performance feature function Z may be less than zero, and then failure may occur with a certain probability. For example, Figure 3.3 shows the failure probability (FP $p(Z<0)$) caused by structural performance degradation. According to equation 3-4, Z is a stochastic variable, which realizations would take the value z, and its density function will be noted $f(z)$. As Z is obtained from the two independent variables R and S, then according to the joint probability theory, the probability density function of Z expresses:

$$f_Z(z)=f_R(r)\times f_S(s), \tag{3-5}$$

where *r* and *s* are the respective values taken by the random variables R and S. Then the joint failure probability corresponding to the probability $P_f(Z<0)$ is depicted by the hatched area in Figure 3.3.

Applying the SORM to the MaxEnt feature yields the PDF $f(z)$, with FP defined by $p(Z<0)$, and failure hazard function (FHF) defined by $\lambda(z)$. Referring to [140], FP and FHF express as follows:

$$p(Z<0)=\int_{Z<0} f_Z(z)dz, \tag{3-6}$$

$$\lambda(z)=\frac{f(z)}{1-p(z<0)}=\frac{f(z)}{R(z)}, \tag{3-7}$$

where the domain of integration $Z<0$ corresponds to the failures domain. In the detection process modeling of early-chatter, the event $Z<0$ means that the chatter feature related to the stresses of the cutting process is higher than necessary, relatively to the stresses in the chatter-free stage. The term necessary means the cutting force and dynamics do not adapt to the strength and characteristics of the

material.

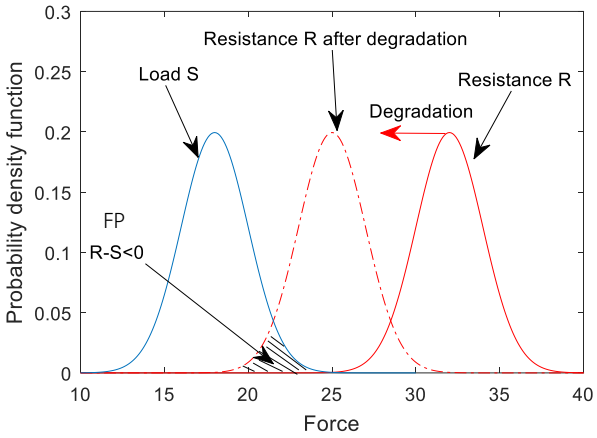


Figure 3. 3 FP caused by structural performance degradation.

The early-chatter can be detected using an indicator from the MaxEnt feature reliability model, starting from a load-resistance-like analysis. The chatter feature extracted from the vibration signal would represent the cutting force produced by the tool in the milling process. The vibration signal contains relevant information on the stochastic chatter process as a result of this analysis. So, one can use the MaxEnt feature as an indicator of this information, a random variable modeled by Z, extracted from the vibration signal packets. Consider z represents the realizations of the random variable Z that describes the stochastic chatter process; one can then estimate the failure rate and calculate the corresponding FHF related to the chatter process. Comparing the obtained FHF to a preset threshold FHF_0 will allow detecting early-chatter.

Let us consider the sampling domain as described in [141] for the system FP. Intuitively, a component operating in a design range close to the limit state may cross this limit to the failure region when subjected to a stochastic process such as the milling process [142].

Therefore, by increasing the number of data sampled on a milling machine, the analysis makes it possible to reach the risk domain included in the sampling domain defined by Harnitz [141] based on the SORM, which would not have been otherwise.

It is necessary to calculate the FHF as the chatter indicator to conduct the early-chatter detection based on the SORM. Thus, the following section illustrates the calculation of the FHF.

(2) Calculation of FHF based on the MaxEnt feature

The signal is segmented suitably into segments or time-windows for detecting the early-chatter reliably and quickly. The time-window width is an essential parameter to trade-off the reliability and the detection delay. Here, the time-window width refers to the number of revolutions in one time-window. The large time-window width leads to a significant detection delay. However, the small time-window width does not provide sufficient information generally. Fortunately, the MaxEnt principle is a significant breakthrough in estimating the probability density function (PDF) [143]. Under the condition of insufficient information, the estimated PDF based on the MaxEnt feature is the optimized and the most unbiased one among

numerous candidates satisfying given conditions.

We introduce the widely used MaxEnt principle with the first four moments for the estimation of the PDF. The estimated PDF is used to calculate the FHF for indicating the appearance of the chatter. The procedure of the estimation of the PDF includes three parts: (a) calculation of moments of sample variables, (b) estimation of the first four moments of the performance feature function, and (c) estimation of the performance feature function.

(a) Calculation of moments of sample variables

In order to estimate the performance feature function using the MaxEnt feature, it is necessary to calculate the moments of chatter features at the chatter-free stage and in the most recent cutting process. Based on the given n samples of a chatter feature x_i, its mean, second, third, and fourth central moment ($\mu_{x_i,1}$, $\mu_{x_i,2}$, $\mu_{x_i,3}$, $\mu_{x_i,4}$) can be estimated as follows:

$$\begin{cases} \hat{\mu}_{x_i,1} = \bar{x}_i = \dfrac{1}{n}\sum_{j=1}^{n} x_{i,j} \\ \hat{\mu}_{x_i,2} = \dfrac{1}{n}\sum_{j=1}^{n} (x_{i,j} - \bar{x}_i)^2 \\ \hat{\mu}_{x_i,3} = \dfrac{1}{n}\sum_{j=1}^{n} (x_{i,j} - \bar{x}_i)^3 \\ \hat{\mu}_{x_i,4} = \dfrac{1}{n}\sum_{j=1}^{n} (x_{i,j} - \bar{x}_i)^4 \end{cases}, \tag{3-8}$$

where $\hat{\mu}_{x_i,1}, \hat{\mu}_{x_i,2}, \hat{\mu}_{x_i,3}, \hat{\mu}_{x_i,4}$ are the estimated value of the $\mu_{x_i,1}, \mu_{x_i,2}, \mu_{x_i,3}, \mu_{x_i,4}$, respectively. $x_{i,j}$ is a sample, i represents ith the chatter feature, j represents the jth sample of the chatter feature x_i. For the early-chatter detection, the first four moments of the chatter feature at the chatter-free stage are estimated as $\hat{\mu}_{x_1,1}, \hat{\mu}_{x_1,2}, \hat{\mu}_{x_1,3}, \hat{\mu}_{x_1,4}$. For the chatter feature in the most recent cutting process, the first four moments are estimated as $\hat{\mu}_{x_2,1}, \hat{\mu}_{x_2,2}, \hat{\mu}_{x_2,3}, \hat{\mu}_{x_2,4}$. Before the early-chatter detection, the first four moments of chatter feature at the chatter-free stage need to be set according to many experiments or experts' knowledge.

(b) Estimation of the first four moments of the performance feature function

According to the estimated first four moments of the chatter features, the first four moments of the performance feature function Z can be estimated based on the Taylor series method. Let us express the performance feature function as $Z = g(x) = g(x_1, x_2, \cdots, x_I)$, where I denotes the number of chatter features. Taylor series expansion of the performance feature function at the mean value expresses as [134]:

$$Z \approx g(\mu) + \sum_{i=1}^{I} g(x_i - \mu_{x_i}) + \frac{1}{2}\sum_{i=1}^{I}\sum_{m=1}^{I} g_{im}(x_i - \mu_{x_i})(x_m - \mu_{x_m}), \tag{3-9}$$

where $\mu = (\mu_{x_1}, \mu_{x_2}, \cdots, \mu_{x_I})$, $g_i = \dfrac{\partial g(\mu)}{\partial x_i}$, $g_{im} = \dfrac{\partial^2 g(\mu)}{\partial x_i \partial x_m}$.

One obtains the first four moments of performance feature function *Z* according to the following equations [144]:

$$\begin{cases} \mu_{Z1} = g(\mu) + \dfrac{1}{2}\sum_{i=1}^{I} g_{ii}\mu_{x_i 2} \\ \mu_{Z2} = \sum_{i=1}^{I} g_i^2 \mu_{x_i 2} + \sum_{i=1}^{I} g_i\, g_{ii}\mu_{x_i 3} + \dfrac{1}{4}\sum_{i}^{I} g_{ii}^2(\mu_{x_i 4} - \mu_{x_i 2}^2) + \dfrac{1}{2}\sum_{i=1}^{I}\sum_{m=1}^{I} g_{ii}^2 \mu_{x_i 2}\mu_{x_m 2} \\ \mu_{Z3} = \sum_{i=1}^{I} g_i^3 \mu_{x_i 3} + \dfrac{3}{2}\sum_{i=1}^{I} g_i^2 g_{ii}(\mu_{x_i 4} - \mu_{x_i 2}^2) + 3\sum_{i=1}^{I}\sum_{m=1}^{I} g_i g_m g_{im}\mu_{x_i 2}\mu_{x_m 2} \\ \mu_{Z4} = \sum_{i=1}^{I} g_i^4 \mu_{x_i 4} + 3\sum_{i=1}^{I}\sum_{m=1}^{I} g_i^2 g_m^2 \mu_{x_i 2}\mu_{x_m 2} \end{cases}. \tag{3-10}$$

(c) Estimation of the performance feature function

For a continuous stochastic variable Z, we assume its PDF as $f(z)$, where *z* represents the realizations of Z. The first moment $\mu_{z,1}$ is calculated as follow:

$$\mu_{z,1} = \int_{\Omega} z f(z) dz, \tag{3-11}$$

where $\Omega = \{-\infty, +\infty\}$ is the integral domain of the stochastic variable *Z*. The second, third, and fourth central moments of the Z are calculated by

$$\mu_{Z,j} = \int_{\Omega} (z - \mu_{Z,1})^j f(z) dz \quad (j = 2,3,4). \tag{3-12}$$

The entropy expresses as

$$H(f(z)) = \int_{\Omega} f(z) \ln f(z) dz. \tag{3-13}$$

Following the MaxEnt feature, the objective is to find the optimal $f(z)$, which maximizes the entropy $H(f(z))$ under some constraints. The MaxEnt principle estimating the optimal PDF $f(z)$ expresses as follows:

$$\max \qquad H(f(z)) = \int_{\Omega} f(z) \ln f(z) dz. \tag{3-14}$$

$$s.t. \quad \begin{cases} \int_{\Omega} f(z) dz = 1 \\ \int_{\Omega} z f(z) dz = \mu_{Z,1} \\ \int_{\Omega} (z - \mu_{Z,1})^j f(z) dz = \mu_{Z,j} \quad (j = 2,3,4) \end{cases} \tag{3-15}$$

We obtain the optimal PDF $f(z)$ by using the Lagrange multipliers method, with the Lagrange function given by

$$L = H(f(z)) - (\lambda_0 - 1)\int_{\Omega} f(z) dz - \lambda_1 \int_{\Omega} z \cdot f(z) dz - \sum_{j=2}^{4} \lambda_j \int_{\Omega} (z - \mu_{Z,1})^j \cdot f(z) dz, \tag{3-16}$$

where $\lambda_0, \lambda_1, \ldots, \lambda_4$ are five Lagrange multipliers relating to the five constraints of (3-15).

To get the maximum entropy, we differentiate *L* with respect to $f(z)$ and then set the resulting expression equal to zero as follows:

$$\frac{\partial L}{\partial f(z)} = 0 \rightarrow -(\ln(f(z))+1) - (\lambda_0 - 1) - \lambda_1 z - \sum_{j=2}^{4} \lambda_j (z - \mu_{Z,1})^j = 0 . \tag{3-17}$$

Equation (3-18) yields the optimal PDF $f(z)$ as it follows:

$$f(z) = e^{-\lambda_0 - \lambda_1 z - \sum_{j=2}^{4} \lambda_j (z-\mu_{Z,1})^j} . \tag{3-18}$$

Inserting (3-18) into the constraints (3-15) allows specifying the five Lagrange multipliers λ_j (j=0,1,2,3,4), and then $f(z)$ is estimated according to (3-18).

After estimating the PDF $f(z)$, one can calculate the FHF using equations (3-6) and (3-7) based on the estimated PDF for early-chatter detection. The calculated FHF enables depicting the state of the milling process.

3.2.3 Decision making

The last section introduced the feature extraction methods, which results are used as input of the decision-making algorithm. The extracted features characterize the state of the monitored process during its operation. The decision is made based on the extracted feature and a preset threshold. However, an improper preset threshold can lead to false/missing alarms and significantly affect the reliability of chatter detection. Thus, setting an optimal threshold is an essential step for obtaining reliable results.

In fault detection based on signal processing, the applied threshold can be categorized into two groups: threshold based on the small probabilities principle (i.e., traditional threshold) and threshold based on two decision risk levels. In this section, we will introduce these two groups of thresholds. To use the proposed adaptive threshold based on two decision risk levels effectively for chatter detection, we choose using the classical sequential probability ratio test (SPRT) [47][46] and a modified SPRT (M-SPRT) [145]. For the details about the classical SPRT and a modified SPRT (M-SPRT), see appendix A or references [47][46][145]

3.2.3.1 Threshold based on the small probabilities principle

In practical probability and statistics, small probability events appear hardly or rarely in a trial, which one defines as the small probability principle. If the small probability events appear in a trial, we will consider the premise condition of events has changed. In early fault detection, the threshold based on the small probabilities principle is the most commonly used method to trigger a decision criterion [93][146]. In practice, the threshold value based on the small probabilities principle is usually set according to the standard deviation σ of the feature at the normal operating stage. For instance, the *3σ* or *6σ* threshold

criterion sets the threshold value as the sum of 3σ or 6σ and the mean of the feature. Once the extracted feature exceeds the preset threshold value, the early fault may occur with a certain probability [93].

In practice, the different measuring instruments or operational environments and conditions influence the feature value. The extracted feature variation can lead to a false or missing alarm or increase the detection delay. Some researchers proposed the adaptive threshold based on the small probabilities principle to deal with this problem according to the noise [147]. The adaptive threshold based on the small probabilities principle can adapt to the noise to ensure fault detection accuracy.

However, the threshold based on the small probabilities principle cannot deal with two decision risk levels caused by the noise. In a noisy environment, the feature has randomness and the normal evolving characteristics as well, so the feature may not cross the value of threshold based on the small probabilities principle in the abnormal process. In this situation, the missing alarm arises. Besides, the feature may cross the threshold value based on the small probabilities principle when the operational process is normal, which will trigger a false alarm. Due to these two decision risk levels, the preset threshold value does not always detect early fault reliably. In other words, the threshold based on the small probabilities principle fails to assess and detect the risk of an early fault accurately.

3.2.3.2 Threshold based on two decision risk levels

In two hypotheses testing theory, two errors (type I α and type II β) were identified based on the probability distribution [47]. The type I error represents the probability of rejecting a true null hypothesis, while the type II error β represents the probability of accepting a false null hypothesis. Therefore, the hypotheses testing theory can deal with the issue mentioned above, so it can be employed to detect early fault reliably. Based on two errors, i.e., decision risks, we can establish the threshold for early-chatter detection.

Unlike the threshold based on the small probabilities principle, the threshold based on two decision risk levels is not affected by the noise. This type of threshold is fixed and only corresponds to two decision risk levels. For varying noise, the threshold based on two decision risk levels can adapt to the noise by adjusting the feature distributions at the fault-free and fault stages. The feature distributions at the fault-free and fault stages are estimated by using the MaxEnt principle mentioned above. In practice, the feature distributions are determined based on a learning process from many experimental results.

It should be noted that the feature distributions need to be determined at the fault-free and fault stages, respectively. However, the threshold based on the small probabilities principle requires only the feature distribution at the fault-free stage. This one may be a disadvantage of the threshold based on two decision risk levels.

According to the two decision risk levels theory, the threshold based on two decision risk levels can be applied to the sequential probability ratio test (SPRT) properly. Appendix A introduces the SPRT and the modified SPRT (M-SPRT) method.

3.3. Early-chatter detection methods

Generally, the chatter signal in the high-speed milling process includes three parts: the periodic components due to the intermittent milling and the rotation of the cutting tool, the chatter component due to the regenerative effect, and the noise component from some uncontrollable influences, such as measuring instrument, in-homogeneous material [108]. The aim of early-chatter detection is to find out the chatter component quickly and reliably.

For the early-chatter component, its amplitude is small. In a noisy environment, it is challenging to detect early-chatter quickly and reliably. To deal with this challenge, we proposed three early-chatter detection methods based on signal processing. The first method is based on the MaxEnt principle and SPRT. This method can detect early-chatter and can be used for remote early-chatter detection. However, this method cannot detect early-chatter relating to the period-2 bifurcation reliably. To address this problem, we develop the second early-chatter detection method based on MSPR and SORM. The second method overcomes the limitation of the first method and can detect early-chatter quickly. Nonetheless, the second method cannot avoid the two decision risk levels due to applying the threshold based on the small probabilities principle. To improve the detection reliability, we proposed the third method based on MSPR and M-SPRT. The proposed three methods are introduced in the following parts.

3.3.1 Early-chatter detection method based on MaxEnt and SPRT

In practice, due to some uncontrollable influences, such as sensor noise and randomness of the parameters, the boundary between the chatter-free and chatter process is fuzzy. The standard deviation of the monitoring may fail to detect early-chatter in this situation.

The MaxEnt principle estimates the PDF from the collected data without specific assumptions under incomplete information. So, the estimated MaxEnt according to the estimated PDF reflects the process characteristics hidden in the data objectively. The SPRT is a cumulative test method, often used to detect the change point in a noisy signal. We can detect early-chatter by combining these two techniques. The flowchart of the proposed early-chatter detection method is given in Figure 3.4.

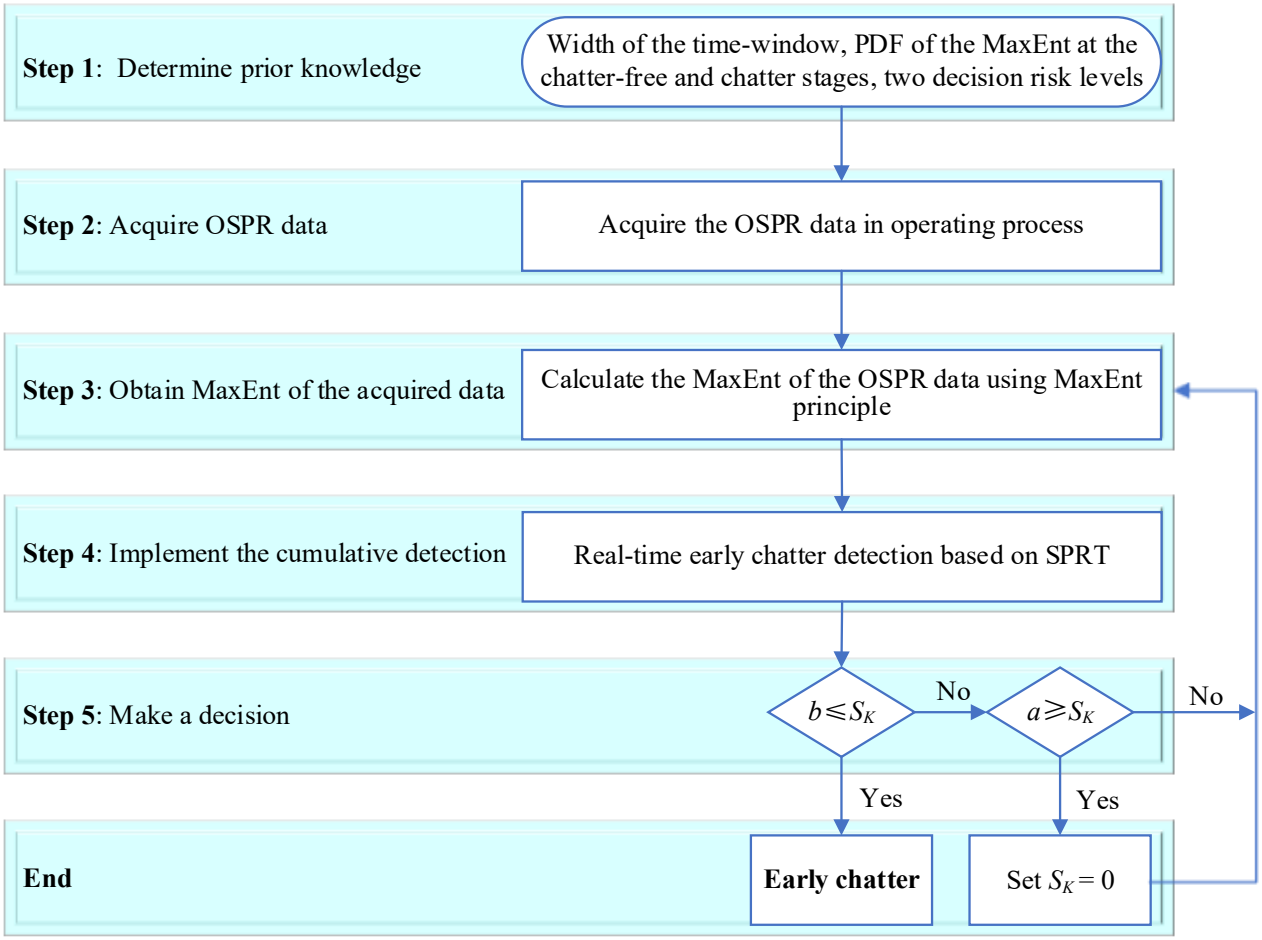


Figure 3. 4 Flowchart of the first proposed early-chatter detection method.

Before detecting early-chatter, some prior knowledge (time-window width for estimating the MaxEnt, the PDF of the MaxEnt at the chatter-free and chatter stages, two decision risk levels) need requiring. The time-window width influences the accuracy of the estimated PDF and the sensitivity of early-chatter detection. The narrow time-window width allows obtaining high sensitivity and a short detection delay. However, the narrower time-window width also leads to a low estimation accuracy of the MaxEnt. Considering the two conflicting aspects and their trade-off, we recommend the width of a time-window lying between 20 to 30 data points.

The setting of the PDF of the MaxEnt at the chatter-free and chatter stages is a crucial step for detecting early-chatter. It is generally set according to a learning process from a large number of experimental studies. Two decision risk levels for SPRT relate directly to the early-chatter detection risk levels or reliability. In the case of a given time-window width, the two decision risk levels cannot be reduced simultaneously. The increase in the time-window width can reduce the two decision risk levels simultaneously. However, the increase in the time-window width would lead to an increase in the detection delay accordingly. The two decision risk levels are set according to the reliability requirements or to the acceptance of risk levels. Traditionally, the two decision risk levels are set between 0.05 and 0.01 in practice. In our case, we set both to 0.01 for this method. The proposed method algorithm consists of five septs as follows:

Step 1: Determine prior knowledge.

Step 2: Acquire OSPR data in real-time and separate the acquired data into some time-windows according to the time-window width.

Step 3: Obtain the MaxEnt of the acquired data per time-window, i.e., evaluate the PDF of the observed data in one time-window using the MaxEnt principle and calculate the MaxEnt according to the evaluated PDF as the feature.

Step 4: Calculate the two threshold values a and b using the two decision risks according to (3-23) and (3-24), and then conduct the SPRT by calculating the CLLR of the obtained MaxEnt.

Step 5: Compare the CLLR with the two threshold values (a and b) and decide as it follows:

- If the CLLR is less than or equal to the threshold value a, the milling process is normal. To avoid the influence of cumulative results on the following test, we reset the CLLR to zero for the following test when it is less than or equal to a. Then, the test procedure returns to step 2 and continue to monitor the dynamic system.
- If the CLLR is more than or equals the threshold value b, it means that an incipient change has occurred and been detected successfully.
- If the CLLR locates in the range of the two threshold values (a and b), the test procedure returns to step 2, and continues to monitor the dynamic system.

3.3.2 Early-chatter detection method based on MSPR and SORM

In this subsection, we propose a MaxEnt feature-based reliability model method, MFRM method, for short, to detect early-chatter based on the MSPR and SORM.

In early-chatter-detection, the sampling frequency, the time-window width, the reference feature, and the threshold FHF_0 need to be set in advance. In order to obtain MSPR data conveniently, the sampling frequency not only satisfies the Nyquist-Shannon sampling theorem but also should be an integral multiple of spindle rotation frequency. The high sampling frequency means more information obtained in one revolution for characterizing the chatter more adequately. However, the high sampling frequency also means a high sampling cost in terms of energy consumption and noise collection. Thus, in practice, the sampling frequency is determined by considering the sampling cost and accuracy in depicting the chatter character. The time-window width is the number of revolutions in one segment and impacts the accuracy and sensitivity of chatter detection. The small time-window width allows reducing the detection delay and obtaining high sensitivity. However, the small time-window width does not reflect the chatter characters and then sufficiently impacts detection accuracy. Considering the two different aspects and their trade-off, we recommend that the time-window width lies between 20 and 30 revolutions. Figure 3-5 illustrates the proposed MFRM method.

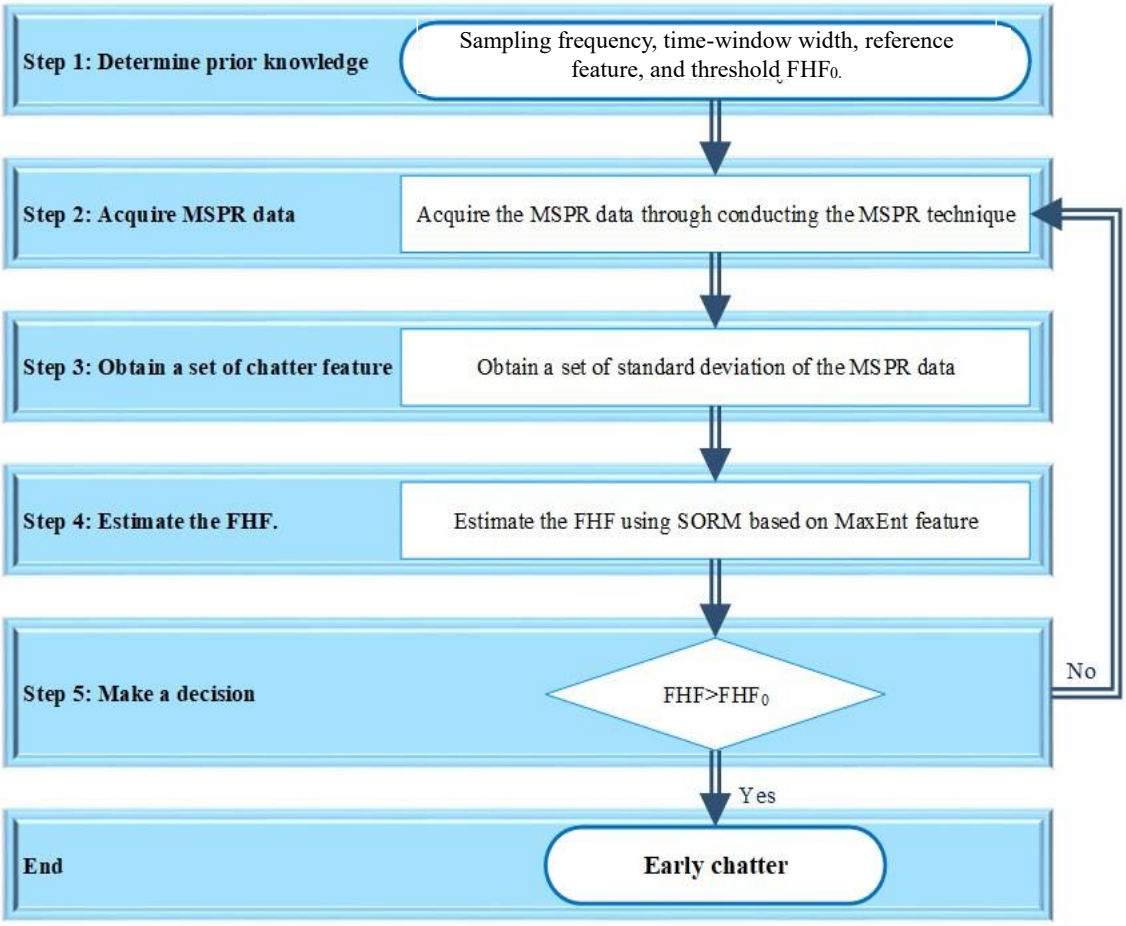


Figure 3. 5 Flowchart of the second proposed method based on SORM and MSPR.

Step 1: Set prior knowledge

Step 2: Collect MSPR data by acquiring the vibration signal and implementing MSPR in real-time.

Step 3: Obtain a set of features by calculating the standard deviation of each line of MSPR data as a feature.

Step 4: Estimate the FHF by using the SORM based on the MaxEnt feature, with the assumption that the standard deviation at the normal stage is the resistance *R* and that the standard deviation of MSPR data in the most recent operating process is the load *S*. The obtained FHF is seen as the chatter indicator.

Step 5: Decide by comparing the obtained FHF and the preset threshold value FHF_0.

- If the FHF is less than or equal to FHF_0, it means the incipient change has not occurred.
- The detection procedure returns to the second step and continues to monitor the dynamics of the system.
- Otherwise, the milling process is out of control, and the early-chatter has appeared and been detected.

For the different machining materials and measuring instruments, the noise of the signal may be different. Less noise may lead to the missing chatter alarm or increases the detection delay because the less noise decreases the standard deviation of the MSPR data. Similarly, more significant noise may cause the fault

chatter alarm. Besides, this method cannot deal with the two decision risks. Aimed at these problems, we develop another early-chatter detection method with less detection delay to adapt to the noise variation.

3.3.3 Early-chatter detection method based on MSPR and adaptive threshold based on two decision risk levels

In this subsection, an reliable detection method of early-chatter is proposed, as Figure 3.6 displays.

For that purpose, some prior knowledge (i.e., the sample number per revolution, time-window width, type I and type II errors, and adaptive threshold based on two decision risk levels) should be set in advance to implement early-chatter detection. The sample number per revolution influences the data acquisition cost and the detection delay. A large sample number per revolution allows employing a short time-window. Then the detection delay can be reduced under the premise of ensuring the estimation quality of the feature. However, a large sample number per revolution may lead to a high data acquisition cost in energy consumption and noise collection. In practice, the sample number per revolution should be set depending on actual applications.

The time-window width is associated with the detection reliability and detection delay. When the time-window width is wide, the feature estimation is more accurate, but the detection delay becomes significant. For early-chatter detection, the early-chatter is related to low-order period-N bifurcations and Hopf bifurcation generally. Based on our experiments and observations, we recommend setting the time-window width between 5 to 10 revolutions depending on the sample number per revolution. The narrow time-window width can be employed when a large sample number per revolution is set. Conversely, a wider time-window width should be chosen when a small sample number per revolution is set.

The type I and type II errors are related to the detection reliability. Thus, in practice, two errors are set according to the reliability requirements or the acceptance of risks. In practice, type I and type II errors are set between 0.05 and 0.01 generally. For this method, we also set both to 0.01.

The adaptive threshold based on two decision risk levels refers to the variable probability distribution of MaxEnt at the chatter-free and chatter stages. The setting of the adaptive threshold based on two decision risk levels at the chatter-free and early-chatter stages is essential for detecting early-chatter. This threshold is set according to a large number of experimental studies or experts' knowledge. The proposed method relies on four steps.

Step 1: Determine the prior knowledge.

Step 2: Acquire MSPR data by acquiring the MSPR data and implementing the MSPR in real-time.

Step 3: Extract the chatter feature by the following three substeps:

a) first, zero-center the acquired MSPR data, which means shifting each set of OSPR data in MSPR data to near zero;
b) then estimate the probability distribution of the zero-centered MSPR data using the MaxEnt

principle;

c) finally, calculate the MaxEnt of the estimated probability distribution as a feature.

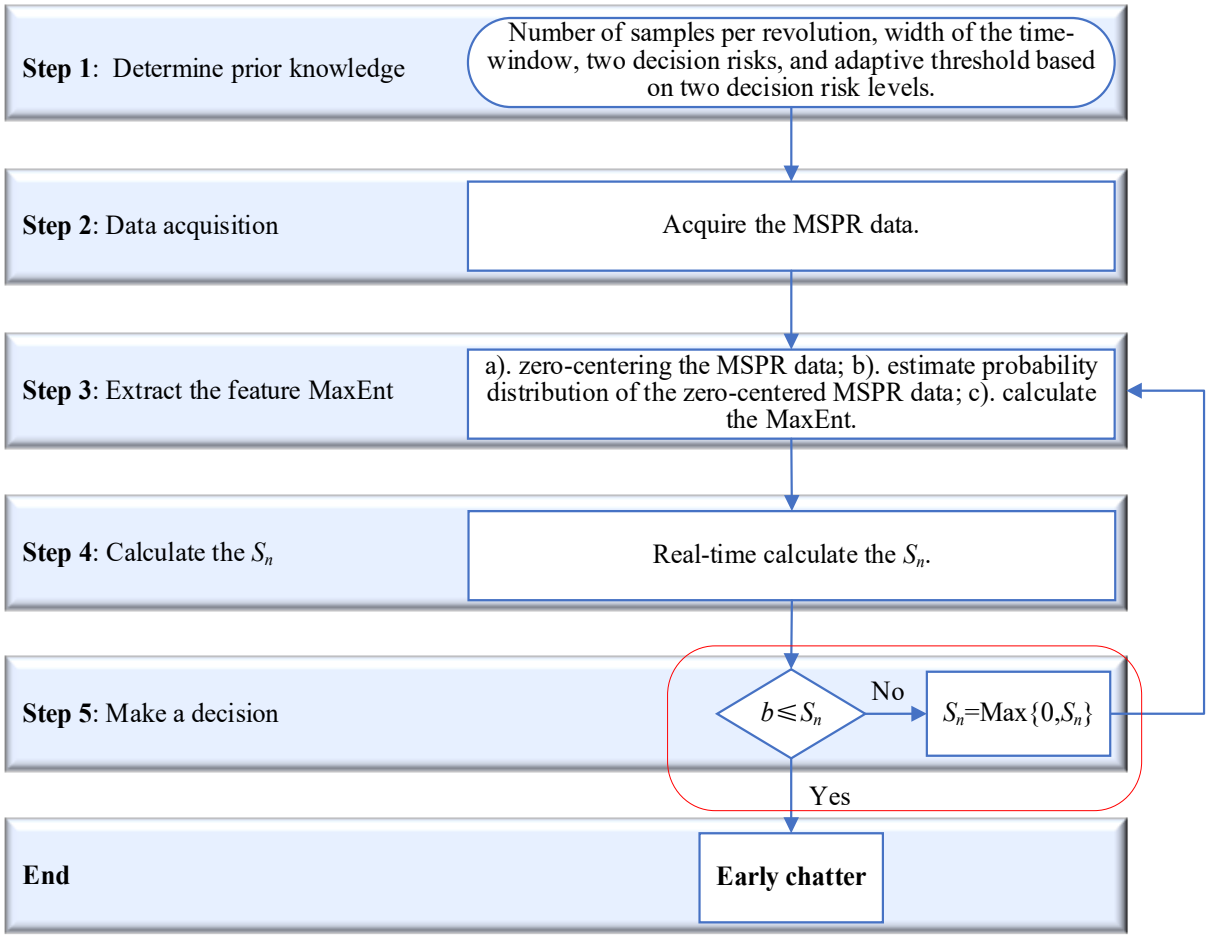


Figure 3. 6 Flowchart of the third proposed detection method based on MSPR and adaptive threshold based on two decision risk levels.

Step 4: Calculate the CLLR S_n based on (3-21) and (3-22).

Step 5: Decide by comparing the CLLR with the upper threshold b.

- If the CLLR S_n is higher than or equal to the threshold b, it means the early-chatter has appeared and been detected.
- Otherwise, the CLLR S_n is reset to Max{0, S_n}. The procedure returns to step 2 and continues to detect early-chatter until the early-chatter is detected or the milling process is stopped.

3.4 Conclusions

In this chapter, we introduce three early-chatter detection methods based on signal processing. The three proposed early-chatter detection methods are based respectively on MaxEnt, MSPR, SORM, SPRT, and M-SPRT.

We present these methods, considering the main steps of general fault detection methods: data acquisition, feature extraction, and decision making. In terms of data acquisition, the three methods use the angular-domain sampled data. Specifically, the first method uses the OSPR data, while the second and third methods use the MSPR data.

In terms of feature extraction, the first and third methods use the MaxEnt principle to estimate the feature MaxEnt, while the second method applies the MFRM to get the feature FHF in the framework of reliability analysis.

In terms of decision making, the first methods employ the threshold based on two decision risk levels and SPRT. The second method utilizes the commonly used threshold based on the small probabilities principle. The third one applies the adaptive threshold based on two decision risk levels and M-SPRT considering the variation of some uncontrollable influences, such as sensor noise and environment.

The three early-chatter detection methods are implemented in a milling process. It should be noted that these three methods also can be used in other dynamic systems, such as the bearing system and power system.

This chapter only introduces three early-chatter detection methods and does not provide their evaluations, which will be given in chapter 5.

Chapter 4 Identification method of period-*N* bifurcations based on SMA and synchronous decomposition

4.1 Introduction

After the early-chatter detection introduced in the last chapter, it is necessary to interpret the chatter phenomenon. The period-*N* bifurcations are common outcomes of the appearance of the early-chatter. In practice, the identification of the period-*N* bifurcations has great significance. For instance, the failure mode or the fault size may be identified by the period-*N* bifurcations [117][116]. In mechanical and electronic systems, the period-*N* bifurcations phenomena frequently appear due to the incorrect setting of the system parameters or system degradation. After the appearance of the period-*N* bifurcations, the signal collected from systems includes subharmonic and inter-harmonic components. The failure mode and the fault size can be identified by investigating the fractal regularities of period-*N* bifurcations or measuring the subharmonic and inter-harmonic [117][116]. Besides, the identification of the period-*N* can be used to obtain some functions in dynamic systems. For instance, the period-*N* bifurcations conditions may be selected to increase the material removal rates in the milling process [13]. To identify the type and size of period-*N* bifurcations, we propose an identification method based on a simple moving average (SMA) and synchronous decomposition.

The proposed identification method includes two steps: period-*N* bifurcations type identification and period-*N* bifurcations size identification. The type identification of period-*N* bifurcations is implemented by using SMA, while the size of period-*N* bifurcations is identified based on the presented synchronous decomposition.

The main contributions of this chapter are summarized as follows:

(I) a method for identifying the type of the period-*N* bifurcations by using SMA.

(II) a method for identifying the period-*N* bifurcations size by using synchronous decomposition.

The rest of the chapter is organized as follows. Section 4.2 introduces the SMA method. Section 4.3 illustrates the synchronous decomposition. Section 4.4 and Section 4.5 show the type and size identification of period-*N* bifurcations by using SMA and synchronous decomposition, respectively. Finally, Section 4.6

lays out some concluding remarks.

4.2 Simple moving average

In the time-domain filtering field, the SMA is a typical finite impulse response filter [23]. The SMA can reduce the random noise and filter the specific frequency and its harmonics effectively with low computational costs. Thus, the SMA is a very suitable tool for real-time fault detection.

The procedure of the SMA relies on calculating the mean or average of subsets of the signal. Given a series of data, $Z=\{z(1), z(2), z(3), \ldots, z(m)\}$, the implementation of the SMA is as follows:

$$y(m)=\frac{1}{M}\sum_{i=0}^{M-1}z(m-i) \quad (M<m), \tag{4-1}$$

where M represents the time-window width. When a new data $z(m+1)$ is collected, we will discard the oldest data $z(m+1-M)$. From this point, we can see that the SMA can be implemented recursively. The recursive formula (4-1) is expressed as follows:

$$y(m+1)=y(m)+\frac{z(m+1)-z(m+1-M)}{M}, \tag{4-2}$$

As mentioned in Chapter 2, SMA has two essential capabilities. One capability is to reduce random noises. The short-term fluctuations are smoothed out, and the longer-term trends or cycles are highlighted. Thus, the SMA can be seen as a low-pass filter. Besides reducing the random noises, the SMA also can filter a specific lower fundamental frequency and its harmonics. In this situation, the SMA can be seen as a special comb filter.

The SMA has only one parameter, i.e., the time-window width. Therefore, the time-window width is an important parameter and determines the performance of the SMA.

The performance of reducing the random noise increases with the increase of the time-window width. If the time-window width is set to the sample number per fundamental period, the SMA can reduce the random noise and filter the specific fundamental frequency and its harmonics.

To illustrate two capabilities of the SMA, we choose two examples. The first example is used to illustrate the capability to reduce random noise. The second example is used to show the capability of both reducing random noise and filtering the specific fundamental frequency and its harmonics.

In the first example, we choose a signal including three components: two low-frequency components and a random white noise component as follows:

$$z(m)=\sin(m\Delta\varphi)+\sin(5m\Delta\varphi)+w(m\Delta\varphi), \tag{4-3}$$

where m is an integer and $\Delta\varphi$ represents equal-angle increment. Two low-frequency components represent signal mode and are seen as the original signal. The third component $w(m\Delta\varphi)$, represents white noise. In practice, the white noise is ubiquitous. Then, we choose the white noise to show the abovementioned

properties of the SMA. The signal to noise ratio (SNR) and the sampling interval $\Delta\varphi$ are set to 5 dB and $1\times10^{-4}\pi$ rad, respectively. Figure 4.1 shows the chosen signal. One can notice that two low-frequency components are buried in a noisy environment. We employ the SMA to reduce the random white noise. To avoid filtering two low-frequency components, we set the time-window width much shorter than the sample number in the periods of two expected components. The sampling interval $\Delta\varphi$ is $1\times10^{-4}\pi$ rad, and the periods of two expected components are 2π and 0.4π, respectively. Consequently, the sample number in periods of the expected components are $20,000$ and 4,000, respectively. In this example, the time-window width is set to 100. Figure 4.2 shows the filtered signal and the original signal. The filtered signal has a little noise and is close to the original signal. Thus, we illustrate the suppression of the noise component significantly by using SMA.

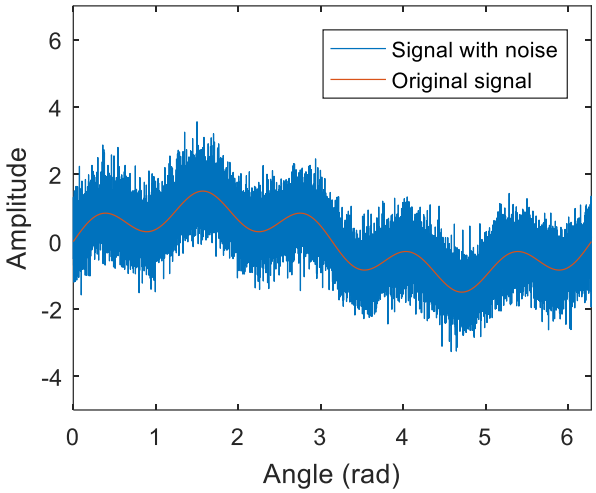


Figure 4. 1 The chosen signal and the original signal.

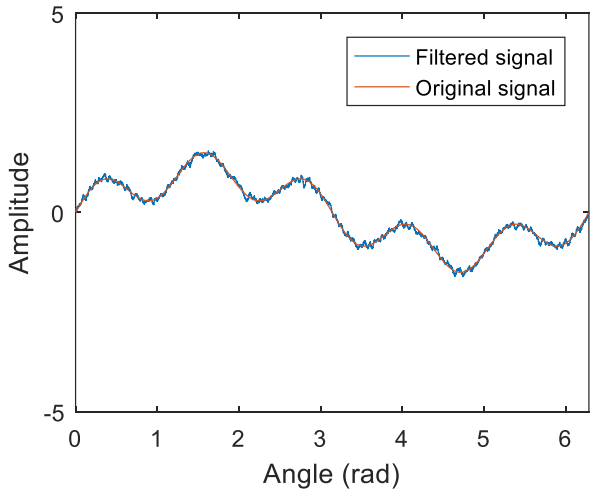


Figure 4. 2 Filtered signal and the original signal.

In the second example, the chosen signal includes four components: three low-frequency components and a random noise component as follows:

$$z(m)=\sin(m\Delta\varphi)+\sin(5m\Delta\varphi)+\sin(10m\Delta\varphi)+w(m\Delta\varphi), \tag{4-4}$$

The first low-frequency component $\sin(m\Delta\varphi)$ represents the original signal. The second and third low-frequency components, i.e., $\sin(5m\Delta\varphi)$ and $\sin(10m\Delta\varphi)$, need to be filtered for highlighting the original one. The third low-frequency component $\sin(10m\Delta\varphi)$ can be seen as a harmonic of the second low-frequency component $\sin(5m\Delta\varphi)$. The fourth component $w(m\Delta\varphi)$ represents a random white noise. The SNR and the sampling interval $\Delta\varphi$ are also set to 5 dB and $1\times10^{-4}\pi$ rad, respectively. Figure 4.3 shows the chosen signal and the chosen signal without the random white noise component $w(m\Delta\varphi)$. From both chosen signals with and without noise component, it is difficult to find the original signal information, i.e., the first low-frequency component $\sin(m\Delta\varphi)$.

The SMA can reduce the random noise and filter the specific frequency and its harmonics. Therefore, we employ the SMA to reduce the random noise and filter the second and third low-frequency components. The time-window width must be set according to the specific frequency and sampling interval for filtering the fundamental frequency and its harmonics. The time-window width needs to equal the sample number *M* per fundamental period:

$$M = \frac{2\pi}{\Delta\varphi}, \tag{4-5}$$

It should be noted that the sample number *M* should be an integer by choosing the sampling interval $\Delta\varphi$. For the second low-frequency component in the chosen signal, the fundamental frequency to be filtered is 5 Hz. As mentioned above, the sampling interval is set to $1\times10^{-4}\pi$. Thus, the time-window width is set to 4,000 samples for filtering the second low-frequency component and its harmonics. Figure 4.4 shows the filtered signal and the original signal. The white noise is reduced significantly. Also, the second low-frequency component and its harmonics are filtered thoroughly. So, it can be seen that the SMA not only suppresses the white noise component significantly but also filters the specific fundamental frequency and its harmonics effectively.

It should be noted that the filtered signal has the same frequency as that of the original signal. However, the amplitudes of the two signals have a slight difference. The averaging of the samples causes this slight difference in the magnitude.

For identifying the period-*N* bifurcations response, the fundamental frequency and its harmonics and random noise need to be filtered to highlight the period-*N* bifurcations components. From the above, we can see that the SMA can filter specific fundamental frequencies and their harmonics and reduce random noise effectively. Thus, the SMA is one potential tool for identifying the type of period-*N* bifurcations.

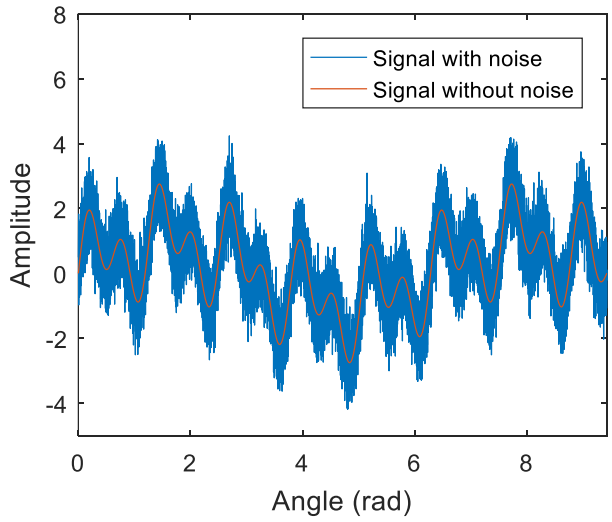


Figure 4. 3 Chosen signal with and without noise.

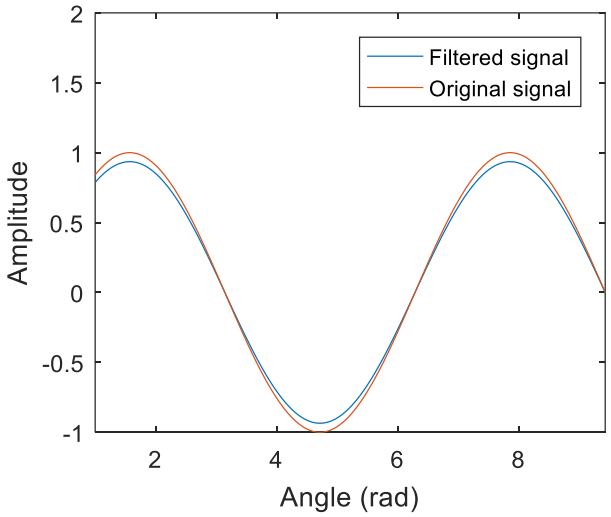


Figure 4. 4 Filtered signal and the original signal.

4.3 Synchronous decomposition

According to the Nyquist–Shannon sampling theorem, the sampling frequency should be greater than twice the highest frequency of the signal to reconstruct a signal fully. The sampling frequency must be far higher than the fundamental frequency to collect the information, including some high-order harmonics. In this case, it is possible to collect some synchronous samples in one revolution in rotating processes.

Therefore, some series of OSPR data can be collected by conducting the OSPR repeatedly. For the stable signal, each data series of OSPR data is constant. However, for the period-*N* bifurcations signal, the fluctuation with specific periods appear in OSPR data. The fluctuation only depends on the property of period-*N* bifurcations. In other words, the period of the OSPR data in the period-*N* bifurcations response is *N* samples. For instance, the period of the OSPR data in the period-2 bifurcation response is two samples. Also, each series of OSPR data represents partial information of the period-*N* bifurcations component. Joining together the partial information provided by each series of OSPR data, we can obtain more accurate

information of period-N bifurcations than that obtained based on only one series of OSPR data. The synchronous decomposition method relates to this aspect.

The synchronous decomposition can be used to decompose the period-N bifurcations signal and extract the period-N bifurcations component. We consider that there are M samples collected in one revolution synchronously. Figure 4.5 illustrates the flowchart of the synchronous decomposition. The procedure of the synchronous decomposition is described as follows:

Step 1: Set two prior parameters: the sample number per revolution, denoted M, and the sample number in OSPR data, denoted n ($n \geq 2$) corresponding to the period-N bifurcations. The sample number M must be an integer, and the sampling interval must be constant. The sample number n in OSPR data is identified in advance and determines the type of period-N bifurcations. The SMA is proposed to identify the type of period-N bifurcations.

Step 2: Acquire MSPR data S by using the MSPR technique. The MSPR technique refers to conduct the OSPR repeatedly. Since the sample number in one revolution is set to M, the acquired MSPR data S is composed of M series of OSPR data.

Step 3. Zero-center each series of OSPR data in MSPR data, and then obtain the period-N bifurcations component S_1.

- Shift the mean of each series of OSPR data in MSPR data to zero. For the stable signal, each series of OSPR data is constant. The fluctuation of the shifted OSPR data only depends on the property of period-N bifurcations. The operation of zero-centering can filter the fundamental frequency and its harmonics.
- For the period-N bifurcations component S_1, if the mean of each series of OSPR data in MSPR data equals zero, the operation of zero-centering will not affect its property. Therefore, the zero-centered OSPR data in MSPR data can depict the period-N bifurcations component S_1. At the end of this section, we will prove that the mean of each series of OSPR data in MSPR data equals zero for the period-N bifurcations component S_1.

Step 4. Subtract the period-N bifurcations component S_1 from the MSPR data S, and then obtain the fundamental component and its harmonics signal S_2.

The MSPR data S includes the period-N bifurcations component S_1 and the fundamental component and its harmonics signal S_2. Thus, the fundamental component and its harmonics signal S_2 can be obtained by subtracting the period-N bifurcations component S_1 from the MSPR data S.

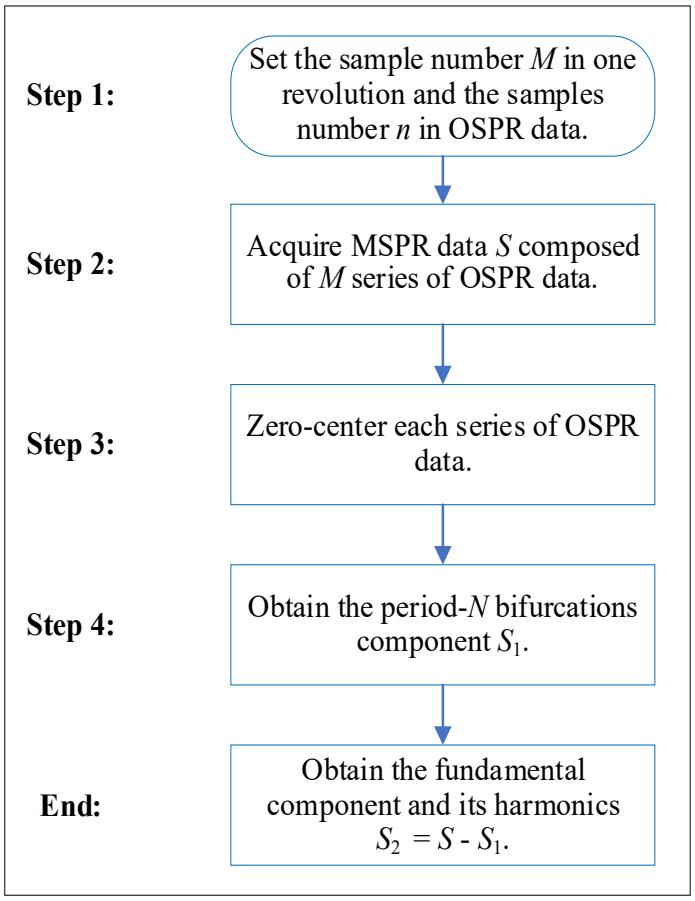


Figure 4. 5 Flowchart of the presented synchronous decomposition.

The period-N bifurcations signal is decomposed into two parts. One part is period-N bifurcations component S_1, including sub-harmonics and inter-harmonics. Another part is the signal S_2, including the fundamental component and its harmonics.

Here is a question: Do all series of zero-centered OSPR data belong to the period-N bifurcations signal? The answer is yes.

Let us validate the presented synchronous decomposition by using two approaches. One approach gives mathematical proof that the mean of OSPR data with n samples corresponding to period-N bifurcations is zero. For the stable signal without noise, the OSPR data is constant. However, for the period-N bifurcations signal without noise, the OSPR data is not constant and has fluctuations. These fluctuations of the OSPR data only depend on the period-N bifurcations. The zero-centered OSPR data will preserve the property of period-N bifurcations when the mean of OSPR data with n samples in the period-N bifurcations component equals zero. Therefore, we validate that the mean of OSPR data with n samples in the period-N bifurcations component equals zero. The other approach is to validate the presented synchronous decomposition using a simulation signal. The effectiveness of synchronous decomposition will be validated if the simulation signal is decomposed accurately.

Without loss of the generality, we assume a period-N bifurcations signal as follows:

$$Y(t) = A_1 \sin(\frac{1}{n} 2\pi ft + \phi_1) + A_2 \sin(\frac{2}{n} 2\pi ft + \phi_2) + \ldots + A_n \sin(\frac{n}{n} 2\pi ft + \phi_n) + A_{n+1} \sin(\frac{n+1}{n} 2\pi ft + \phi_{n+1})$$
$$+ A_{n+2} \sin(\frac{n+2}{n} 2\pi ft + \phi_{n+2}) + \ldots + A_{(k-1)n+i} \sin(\frac{(k-1)n+i}{n} 2\pi ft + \phi_{(k-1)n+i}) + \ldots \quad ,$$
$$+ A_{kn} \sin(\frac{kn}{n} 2\pi ft + \phi_{kn}), \quad (k = 1,2,3,...; i = 1,2,3,...; n \subset 2,3,4,...)$$

(4-7)

where *f* represents fundamental frequency, *n* represents the type of the period-*N* bifurcations and the sample number in OSPR data. The $\frac{f}{n}$ is the sub-harmonic frequency. For the period-*N* bifurcations, *n* is greater than or equal to two. A_{kn} and ϕ_{kn} are the amplitude and phase of the kn^{th} sinusoidal, respectively. *k* represents the *k* order harmonic. The period-*N* bifurcations signal has the smallest period of nT, T represents the fundamental period and equals $\frac{1}{f}$.

The signal $Y(t)$ can be separated into two parts. One part is the fundamental component, and its harmonics component is represented by sinusoidal $A_{kn} \sin(2k\pi ft + \varphi_{kn})$. The other part defined by $A_{(k-1)n+i} \sin(\frac{(k-1)n+i}{n} 2\pi ft + \phi_{(k-1)n+i})$ is the period-*N* bifurcations component. For the sinusoidal signal $A_{kn} \sin(2k\pi ft + \varphi_{kn})$ representing the fundamental component and its harmonics component, the OSPR data is constant. The fluctuations of the OSPR data only depend on the sinusoidal signal $A_{(k-1)n+i} \sin(\frac{(k-1)n+i}{n} 2\pi ft + \varphi_{(k-1)n+i})$ representing the period-*N* bifurcations component. The proof will be obtained if the mean of *n* samples in OSPR data related to each component $A_{(k-1)n+i} \sin(\frac{(k-1)n+i}{n} 2\pi ft + \varphi_{(k-1)n+i})$ equals zero.

For each sinusoidal signal $A_{(k-1)n+i} \sin(\frac{(k-1)n+i}{n} 2\pi ft + \varphi_{(k-1)n+i})$ representing the period-*N* bifurcations component, the *n* samples in OSPR data can be noted $Y_{(k-1)n+i}\{y^0_{(k-1)n+i}, y^1_{(k-1)n+i}, y^2_{(k-1)n+i}, ..., y^{n-1}_{(k-1)n+i}\}$ as follows;

$$Y_{(k-1)n+i}=\left\{\begin{array}{c} A_{(k-1)n+i}\sin(\frac{(k-1)n+i}{n}2\pi ft+\varphi_{(k-1)n+i}) \\ A_{(k-1)n+i}\sin(\frac{(k-1)n+i}{n}2\pi ft+\varphi_{(k-1)n+i}+\frac{1}{n}((k-1)n+i)2\pi) \\ A_{(k-1)n+i}\sin(\frac{(k-1)n+i}{n}2\pi ft+\varphi_{(k-1)n+i}+\frac{2}{n}((k-1)n+i)2\pi) \\ \vdots \\ A_{(k-1)n+i}\sin(\frac{(k-1)n+i}{n}2\pi ft+\varphi_{(k-1)n+i}+\frac{n-1}{n}((k-1)n+i)2\pi) \end{array}\right\}^{T}. \quad (4\text{-}8)$$

For convenience, we assume:

$$A=A_{(k-1)n+i}\ , \quad (4\text{-}9a)$$

$$K=(k-1)n+i\,, \quad (4\text{-}9b)$$

$$x=\frac{(k-1)n+i}{n}2\pi ft+\phi_{(k-1)n+i}\,. \quad (4\text{-}9c)$$

Next, we give the proof that the mean of *n* OSPR samples Y_K equals zero. The mean of the Y_K can be expressed as follows:

$$E(Y_K)=\frac{A}{n}(\sin(x)+\sin(x+\frac{1}{n}2K\pi)+\sin(x+\frac{2}{n}2K\pi)+\ldots+\sin(x+\frac{n-1}{n}2K\pi))\,. \quad (4\text{-}10)$$

Let us recall the sum-to-product and product-to-sum formula as follows:

$$\sin(a)+\sin(b)=2\sin(\frac{a+b}{2})\cos(\frac{a-b}{2})\,, \quad (4\text{-}11a)$$

$$2\sin(a)\cos(b)=\sin(a+b)+\sin(a-b)\,, \quad (4\text{-}11b)$$

If the sample number in OSPR data *n* is even, according to the sum-to-product (4-11a), the mean $E(Y_K)$ (4-10) is rewritten as follows:

$$\begin{aligned} E(Y_K)&=\frac{A}{n}(\sin(x)+\sin(x+\frac{n-1}{n}2K\pi)+\sin(x+\frac{1}{n}2K\pi)+\sin(x+\frac{n-2}{n}2K\pi)+ \\ &\ldots+\sin(x+\frac{n-2}{2n}2K\pi)+\sin(x+\frac{n}{2n}2K\pi)) \\ &=\frac{1}{n}(2\sin(x+\frac{n-1}{n}K\pi)\cos(-\frac{n-1}{n}K\pi)+2\sin(x+\frac{n-1}{n}K\pi)\cos(-\frac{n-3}{n}K\pi)+ \\ &\ldots+2\sin(x+\frac{n-1}{n}K\pi)\cos(x+\frac{1}{n}K\pi)) \\ &=\frac{1}{n}2\sin(x+\frac{n-1}{n}K\pi)(\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+ \\ &\ldots+\cos(\frac{3}{n}K\pi)+\cos(\frac{1}{n}K\pi)) \end{aligned}\,, \quad (4\text{-}12)$$

Considering the property of (4-12), we multiply $2\sin(\frac{1}{n}K\pi)$ first and then divide $2\sin(\frac{1}{n}K\pi)$. Due to

$K=(k-1)n+i$, the $2\sin(\frac{1}{n}K\pi)$ does not equal zero and can be a denominator. According to the product-to-sum formula (4-11b), the mean $E(Y_K)$ (4-12) is rewritten as follows:

$$\begin{aligned}
E(Y_K) &= \frac{2A}{n}\sin(x+\frac{n-1}{n}K\pi) \\
&\cdot \frac{2\sin(\frac{1}{n}K\pi)(\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+\ldots+\cos(\frac{1}{n}K\pi))}{2\sin(\frac{1}{N}K\pi)} \\
&= \frac{2A}{n}\sin(x+\frac{n-1}{n}K\pi) \\
&\cdot \frac{2\sin(\frac{1}{n}K\pi)\cos(\frac{n-1}{n}K\pi)+2\sin(\frac{1}{n}K\pi)\cos(\frac{n-3}{n}K\pi)+\ldots+2\sin(\frac{1}{n}K\pi)\cos(\frac{1}{n}K\pi))}{2\sin(\frac{1}{n}K\pi)} \\
&= \frac{2A}{n}\sin(x+\frac{n-1}{n}K\pi) \\
&\cdot \frac{\sin(\frac{n}{n}K\pi)-\sin(\frac{n-2}{n}K\pi)+\sin(\frac{n-2}{n}K\pi)-\sin(\frac{n-4}{n}K\pi)+\ldots+\sin(\frac{2}{n}K\pi)}{2\sin(\frac{1}{n}K\pi)} \\
&= \frac{2A}{n}\sin(x+\frac{n-1}{n}K\pi)\frac{\sin(\frac{n}{n}K\pi)}{2\sin(\frac{1}{n}K\pi)}
\end{aligned} \tag{4-13}$$

Due to $K=(k-1)n+i$, the $\sin(\frac{n}{n}K\pi)=0$. Thus, the mean $E(Y_K)$ equals zero when the number *n* of samples in OSPR data is even.

If the number *n* of samples in OSPR data is odd, the mean $E(Y_K)$ can be rewritten by adjusting the order of sinusoidal as follows:

$$\begin{aligned}
E(Y_K) = \frac{A}{n}(&\sin(x)+\sin(x+\frac{n-1}{n}2K\pi)+\sin(x+\frac{1}{n}2K\pi)+\sin(x+\frac{n-2}{n}2K\pi)+ \\
&\ldots+\sin(x+\frac{n-1}{2n}2K\pi))
\end{aligned}, \tag{4-14}$$

According to the sum-to-product (4-11a), the mean $E(Y_K)$ (4-14) can be rewritten as follows:

$$
\begin{aligned}
E(Y_K) &= \frac{2A}{n}\{\sin(x+\frac{n-1}{n}K\pi)\cos(-\frac{n-1}{n}K\pi)+2\sin(x+\frac{n-1}{n}K\pi)\cos(-\frac{n-3}{n}K\pi)+ \\
&\quad \ldots+\sin(x+\frac{n-1}{n}K\pi)\} \\
&= \frac{2A}{n}\sin(x+\frac{n-1}{n}K\pi)\{\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+ \\
&\quad \ldots+\cos(\frac{2}{n}K\pi)+\frac{1}{2}\}
\end{aligned}
\quad , (4\text{-}15)
$$

The mean $E(Y_K)$ will be equal to zero if the following equality (4-16) is established.

$$
\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+\ldots+\cos(\frac{2}{n}K\pi)=-\frac{1}{2}. \quad (4\text{-}16)
$$

We assume:

$$
B=\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+\ldots+\cos(\frac{2}{n}K\pi). \quad (4\text{-}17)
$$

Considering the characteristic of *B* (4-17), let us multiply $2\sin(\frac{1}{n}K\pi)$ to the right of the equation (4-17) and then divide $2\sin(\frac{1}{n}K\pi)$ as follows:

$$
B=\frac{2\sin(\frac{1}{n}K\pi)(\cos(\frac{n-1}{n}K\pi)+\cos(\frac{n-3}{n}K\pi)+\cos(\frac{n-5}{n}K\pi)+\ldots+\cos(\frac{2}{n}K\pi))}{2\sin(\frac{1}{n}K\pi)}. \quad (4\text{-}18)
$$

As mentioned above, due to $K=(k-1)n+i$, the $2\sin(\frac{1}{n}K\pi)$ does not equal zero and can be a denominator. According to the product-to-sum (4-11b), the (4-18) is derived as follows:

$$
\begin{aligned}
B &= \frac{\sin(\frac{n}{n}K\pi)-\sin(\frac{n-2}{n}K\pi)+\sin(\frac{n-2}{n}K\pi)-\sin(\frac{n-4}{n}K\pi)+\cdots+\sin(\frac{3}{n}K\pi)-\sin(\frac{1}{n}K\pi)}{2\sin(\frac{1}{N}K\pi)} \\
&= \frac{\sin(K\pi)-\sin(\frac{1}{n}K\pi)}{2\sin(\frac{1}{n}K\pi)} \\
&= \frac{-\sin(\frac{1}{n}K\pi)}{2\sin(\frac{1}{n}K\pi)} \\
&= -\frac{1}{2}
\end{aligned}
\quad . (4\text{-}19)
$$

The value of *B* equals $-\frac{1}{2}$. Thus, the mean $E(Y_K)$ equals zero when the number *n* of samples in OSPR

data is odd.

From the above, we can see that the mean $E(Y_K)$ equals zero for all sample number $n(n \geq 2)$ in OSPR data. The proof shows that the procedure of zero-centering each series of OSPR data does not impact the property of the period-N bifurcations. Therefore, the presented synchronous decomposition can decompose the period-N bifurcations response accurately.

Next, we use an example to validate the effectiveness of the presented synchronous decomposition. The example employs a mixed-signal as follows:

$$y(t) = 0.2\sin(100\Delta\varphi) + \sin(200\Delta\varphi). \tag{4-20}$$

The mixed-signal includes two components. The frequencies of the two components are $\frac{50}{\pi}$ Hz and $\frac{100}{\pi}$ Hz, respectively. We regard the mixed-signal as a period-2 bifurcation response. The $\frac{100}{\pi}$ Hz is seen as a fundamental frequency. The $\frac{50}{\pi}$ Hz is seen as a sub-harmonic frequency. We set the sampling frequency to $\frac{40}{\pi}$ kHz. Figure 4.6, 4.7, and 4.8 show the mixed-signal, the first component, and the second component, respectively.

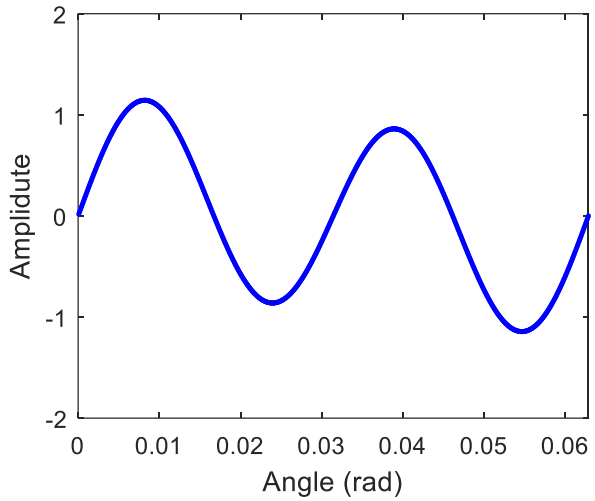


Figure 4. 6 Mixed-signal.

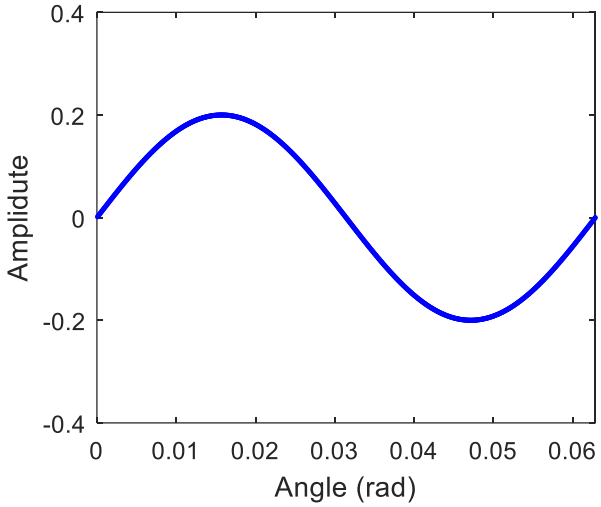


Figure 4. 7 First component with a frequency of $\frac{50}{\pi}$ Hz.

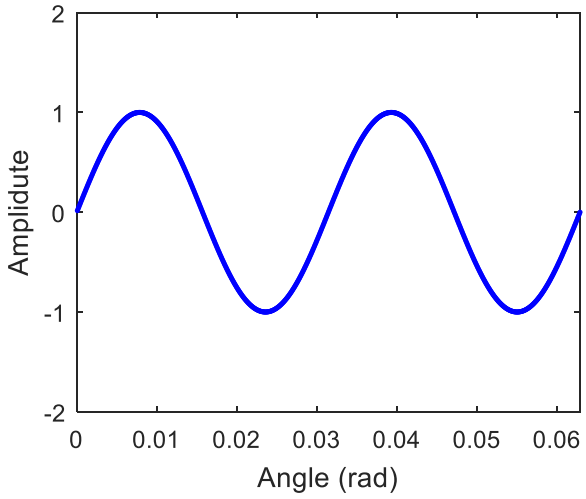


Figure 4. 8 Second component with a frequency of $\frac{100}{\pi}$ Hz.

Since the mixed signal is a linear combination of sinusoidal components, some signal processing methods can identify or decompose the mixed signal. However, the existing signal processing methods fail to decompose the mixed signal or identify the period-N bifurcations accurately when the time-window is short. For instance, when the time-window is one period of the mixed-signal, i.e., $0.01\,\pi$ s, the Fast Fourier transform (FFT) fails to identify the period-2 bifurcation accurately. Figure 4.9 shows the results of the FFT. The frequency resolution is 50 Hz since the time-window is $0.01\,\pi$ s. Due to the rough frequency resolution of the FFT, it is difficult to identify the mixed-signal.

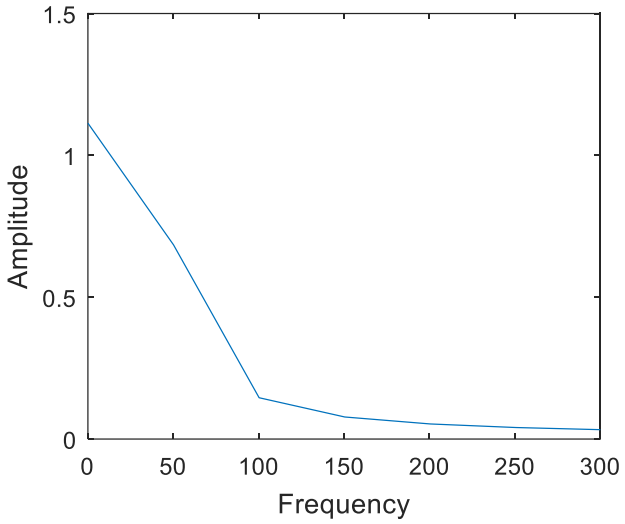


Figure 4. 9 FFT results with a frequency resolution of 50 Hz.

The synchronous decomposition can decompose the mixed-signal accurately. The sampling frequency and the fundamental frequencies are $\frac{40}{\pi}$ kHz and $\frac{100}{\pi}$ Hz, respectively. Thus, the sample number M is 400 in the fundamental period. One can collect 400 series of OSPR data. The time-window width is set to 800 for period-2 bifurcation. Therefore, each series of OSPR data has two samples.

The synchronous decomposition is used to decompose the mixed signal. Figure 4.10 and 4.11 show two decomposed components, respectively. Two decomposed components are quite close to two original components shown in Figures 4.7 and 4.8. Figure 4.12 and 4.13 show the decomposition error of two components. It should be noted that the scale is 10^{-15}. The decomposition error is quite small and can be seen as a computational error. Therefore, the presented synchronous decomposition method decomposes the mixed-signal accurately. Besides, the presented synchronous decomposition method has a low computational cost. In this example, the computational cost is about 0.0029 s.

It should be noted that the sample number M in one revolution does not impact the decomposition accuracy of the synchronous decomposition. However, a significant sample number M in one revolution can depict high-order harmonics and show the signal clearly.

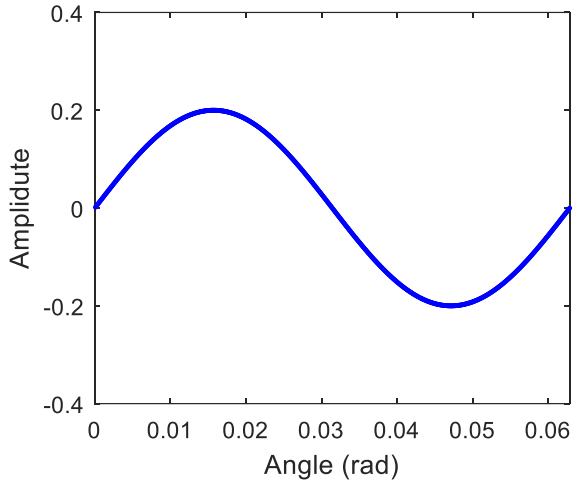


Figure 4. 10 First decomposed component with a frequency of $\frac{50}{\pi}$ Hz.

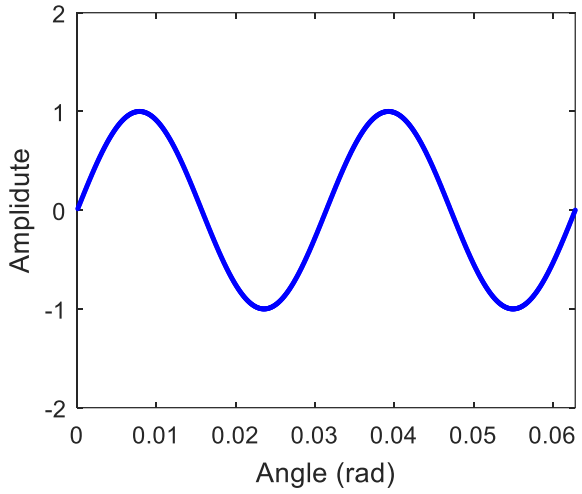


Figure 4. 11 Second decomposed component with a frequency of $\frac{100}{\pi}$ Hz.

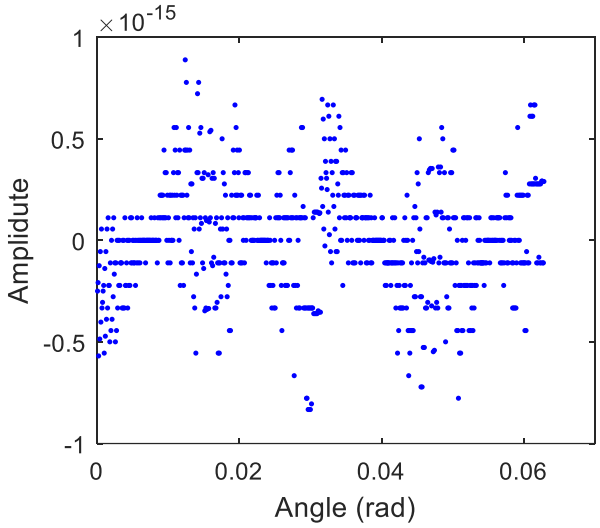


Figure 4. 12 Decomposed error of the first component with a frequency of $\frac{50}{\pi}$ Hz.

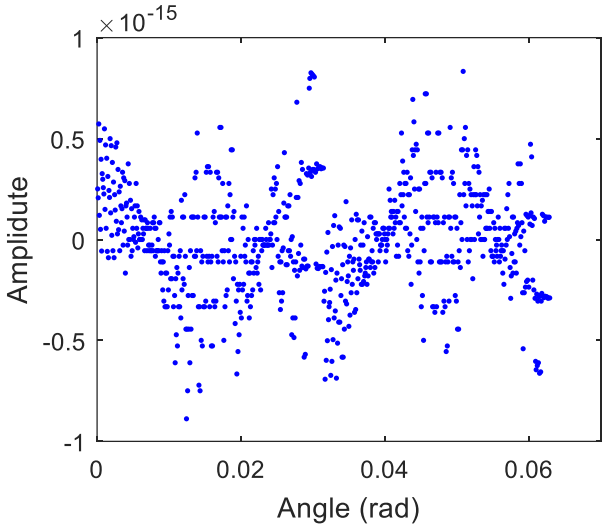


Figure 4. 13 Decomposed error of the second component with a frequency of $\frac{100}{\pi}$ Hz.

4.4 Period-*N* bifurcations type identification based on SMA

In some dynamic systems, such as power supply systems and machining systems, the appearance of the period-*N* bifurcations refers to the abnormal operation or failure. Thus, the type identification of period-*N* bifurcations has significance in fault diagnosis [115][116][87]. Besides, the machining quality obtained under conditions of period-*N* bifurcations is better than that obtained under conditions of Hopf bifurcation. To improve the milling process's productivity, some researchers proposed to mill the work-piece under the conditions of the period-*N* bifurcations [12][13]. Therefore, the type identification of period-*N* bifurcations has significance in the milling process.

This section proposes a type identification method of period-*N* bifurcations through an SMA. The proposed method uses the SMA and standard deviation (SD) to identify the type of period-*N* bifurcations. As mentioned in chapter 2, the sub-harmonic sampling strategy and response integral quantity method have been applied to identify the type of period-*N* bifurcations [87][66]. However, the sub-harmonic sampling may fail to identify period-*N* bifurcations because the sub-harmonic sampling data includes partial information of period-*N* bifurcations. In a noisy environment, the response integral quantity method is not implementable easily. Also, in [87][66], the authors did not consider the noise surrounding the signal. In practice, the noise is inevitable. The SMA can reduce the random noise and filter the specific frequency and its harmonics. Thus, the use of SMA is a suitable tool to identify the type of period-*N* bifurcations.

Figure 4.14 illustrates the flowchart of the proposed type identification method of period-*N* bifurcations. The procedure of the type identification method can be illustrated as follows:

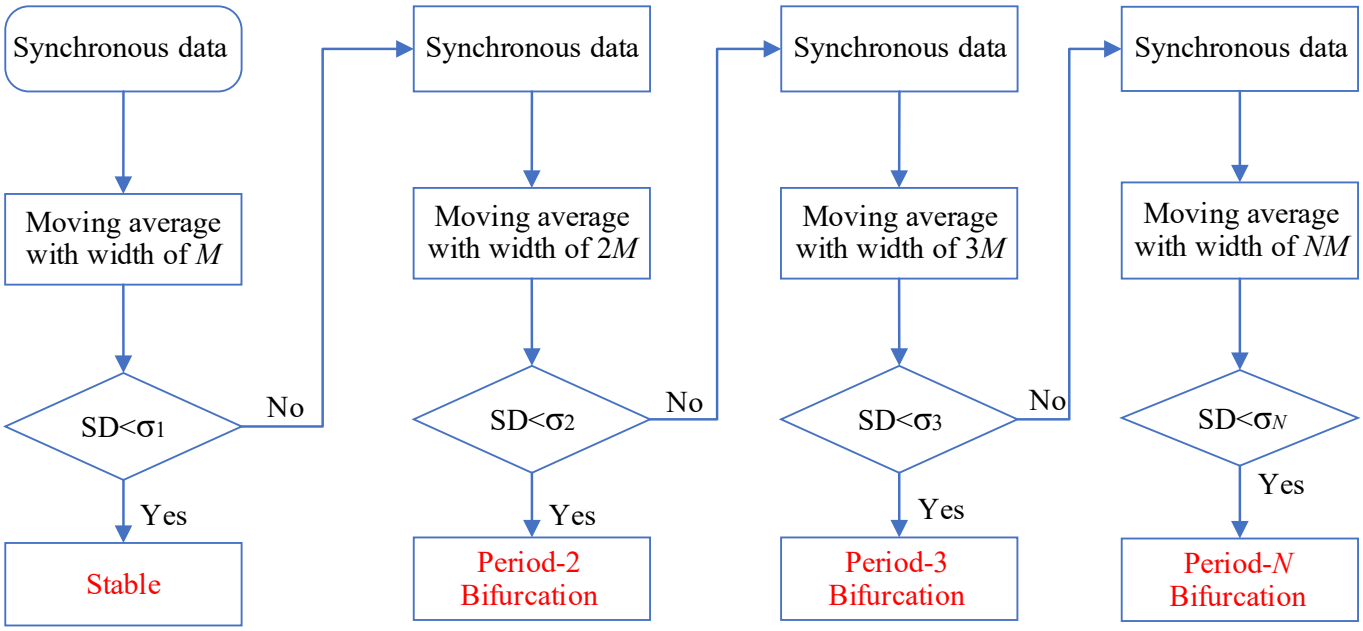


Figure 4. 14 Flowchart of the proposed type identification method of period-*N* bifurcations.

Step 1. Set the sample number *M* in one fundamental period (or excitation period) and a parameter n=1.

Step 2. Acquire synchronous data S.

Step 3. Set the time-window width to *nM*.

Step 4. Use the SMA to process the acquired synchronous signal.

Step 5. Calculate the standard deviation (SD) of the filtered signal.

Step 6. Compare the calculated SD with the preset threshold value σ_n. If the SD is less than the σ_n. The process is the period-N bifurcations (period-1 bifurcation response is seen as a stable process). Otherwise, move the procedure to the next step.

Step 7. Reset $n = n+1$ and return the procedure to step 3 for continuing to identify the type of period-*N* bifurcations.

The procedure will be stopped when the period-*N* bifurcations response is identified.

4.5 Period-*N* bifurcations size identification based on synchronous decomposition

In fault diagnosis, the size of the period-*N* bifurcations can relate to the fault size. Therefore, the size identification of the period-*N* bifurcations can be used to identify the fault size. To identify the period-*N* bifurcations size accurately, we present a size identification method of period-*N* bifurcations based on the presented synchronous decomposition. The method is based on the results of the method for type identification of period-*N* bifurcations we proposed in the previous section. Therefore, the type of period-*N* bifurcations should be identified in advance to implement the proposed size identification method of period-*N* bifurcations. Figure 4.15 shows the flowchart of the presented size identification method of period-*N* bifurcations. The procedure of the method can be illustrated as follows:

Step 1. Set the time-window width to *nM* according to the identification results in the first part.

Step 2. Decompose the synchronous data S by using the presented synchronous decomposition, and obtain the period-*N* bifurcations component S1.

Step 3. Calculate the ratio R between the SD of the period-*N* bifurcations component S1 and synchronous data S.

Step 4. Indicate the size of the period-*N* bifurcations using the calculated ratio R.

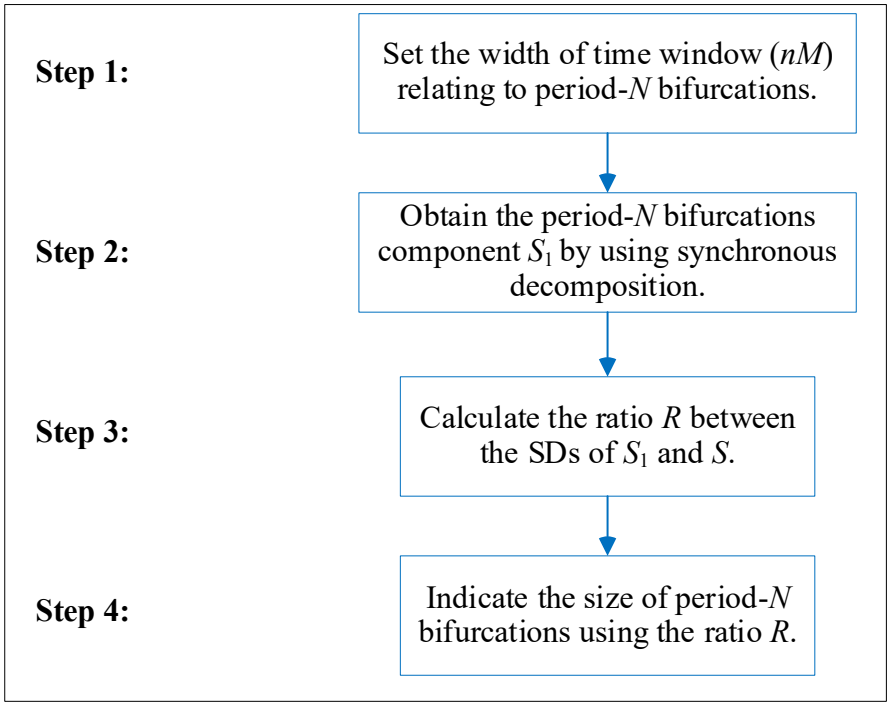


Figure 4. 15 Flowchart of the size identification method of period-N bifurcations.

4.6 Conclusion

This chapter proposes an identification method for identifying the type and size of period-*N* bifurcations. The proposed identification method of period-*N* bifurcations includes two parts: type identification of period-*N* bifurcations and size identification of period-*N* bifurcations. An SMA allows identifying the type of period-*N* bifurcations. The SMA can reduce the random noise and filter the specific fundamental frequency and its harmonics. Thus, the period-*N* bifurcations component is highlighted, which allows identifying the type of period-*N* bifurcations. Then the synchronous decomposition is implemented to identify the size of period-*N* bifurcations. Besides, proof of the synchronous decomposition is provided. The synchronous decomposition allows extracting the period-*N* bifurcations component from synchronous data. Based on the period-*N* bifurcations component and the synchronous data, we can identify the period-*N* bifurcations size.

Chapter 5 Evaluation of early-chatter detection methods and period-*N* bifurcations identification method

5.1 Introduction

The last two chapters introduced three early-chatter detection methods and a period-*N* bifurcations identification method. This chapter aims to assess the performance of these methods using a milling process case study. This case is a typical machining process commonly used in the manufacturing fields. Due to the property of the intermitted cutting, milling is a complex nonlinear process. In normal cutting conditions, the milling process is stable. The frequencies of the stable milling signal include spindle revolution frequency, tooth passing frequency, and their corresponding harmonics. Whereas, in the abnormal cutting process, the milling process is unstable. The unstable milling signal, besides the spindle revolution frequency, tooth passing frequency, and corresponding harmonics, also involves chatter frequencies. Therefore, we choose the milling process to evaluate the performance of the proposed early-chatter detection methods.

Also, we choose the milling process as a case study to evaluate the performance of the proposed period-*N* bifurcations identification method. For the purpose of diagnosis or control of the chatter, after the early-chatter detection, it is necessary to identify chatter [13][12]. In general, a local phenomenon of early-chatter may also relate to a bifurcation phenomenon that can get worse and turn to chaos with the machine dynamics when it becomes a severe chatter. In a milling process, the period-*N* bifurcations and the Hopf bifurcation may appear. The period-*N* bifurcations appear more often in the milling process due to the impact of improper machining parameters [13]. Thus, the identification of the period-*N* bifurcations should allow recognizing the inappropriate machining parameters. Also, the identification of the period-*N* bifurcations has practical significance for improving productivity in a milling process. According to the works reported in [13][12], some milling processes under the period-*N* bifurcation conditions have higher machining quality than those under the Hopf bifurcation condition. Some machining parameters corresponding to the period-*N* bifurcations can be used to improve productivity and ensure the required machining quality. To this end, the period-*N* bifurcations need to be identified during the milling process.

We organize the rest of the chapter as follows. Section 5.2 introduces the chosen milling case for validating

the proposed methods. Section 5.3 is devoted to approving the proposed early-chatter detection methods. Then Section 5.4 assesses the proposed period-*N* bifurcations identification method, and lastly, Section 5.5 draws concluding remarks.

5.2 Milling process models for evaluating the proposed methods

In this section, we introduce two milling process models for evaluating the performance of the proposed methods. In the milling process, improper machining parameters or machine faults can lead to the appearance of chatter. Thus, to evaluate the proposed methods, we choose two milling models: one with a machine fault represented by the clearance of the bearing installed in the spindle, and the other without machine fault. The milling process model with bearing clearance fault is used to simulate the chatter caused by machine fault and validate the proposed early-chatter detection methods. The milling process model without machine faults is used to simulate the chatter caused by improper machining parameters and prove the effectiveness of the period-*N* bifurcations identification method.

5.2.1 Milling process model with bearing clearance fault

To assess the three early-chatter detection methods we proposed in chapter 3, we choose the high-speed milling simulation model with the bearing clearance fault introduced in [1] to generate the vibration signal. Figure 5.1 depicts the high-speed milling simulation model. The cutting tool is equivalent to a two-degree-of-freedom (2-DOF) dynamic system in the feed (*x*-) and cross-feed (*y*-) directions. The work-piece is assumed rigid. The variable stiffness represents the effect of the bearing clearance fault on the dynamic system [1][148].

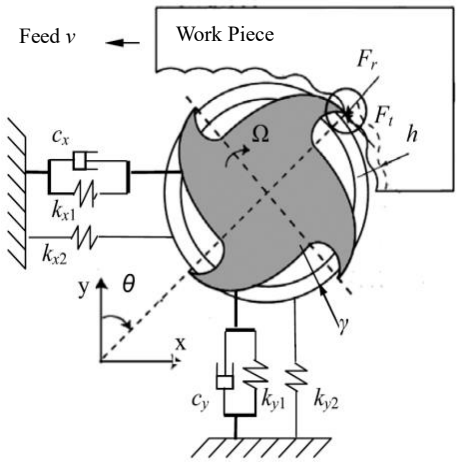


Figure 5. 1 Dynamic model of the milling process with bearing clearance fault [1].

The 2-DOF dynamic milling system is excited by two variable cutting forces in *x*-direction and *y*-direction. According to D'Alambert principle, its governing dynamic equations are:

$$\begin{cases} m_x\ddot{x} + c_x\dot{x} + k_{x1}x + \delta(r-\gamma)k_{x2}(r-\gamma)\cos(\theta) = F_x(t) \\ m_y\ddot{y} + c_y\dot{y} + k_{y1}y + \delta(r-\gamma)k_{y2}(r-\gamma)\cos(\theta) = F_y(t) \end{cases}, \tag{5-1}$$

where x, $\dot{x}$, $\ddot{x}$, y, $\dot{y}$ and $\ddot{y}$ are the displacement, the velocity, and the acceleration of the cutting tool tip in x-direction and y-direction, respectively. Furthermore, in the model, m_x, m_y, c_x, c_y, $k_{x_1}+\delta(r-\gamma)k_{x_2}$, and $k_{y_1}+\delta(r-\gamma)k_{y_2}$ are parameters of mass, damping, and stiffness in x-direction and y-direction, respectively, with a polar variable $r=\sqrt{x^2+y^2}$, $\cos(\theta)=x/r$, $\sin(\theta)=y/r$, and θ is the rotation angle of the spindle. The variable γ is the radial clearance of the bearing. The variable $\delta(\cdot)$ is a switch function with $\delta(\Delta)=1$, if $\Delta\geq 0$, and $\delta(\Delta)=0$, if $\Delta<0$, where Δ represents a variable. Table 5.1 displays the model parameters of the dynamic system. $F_x(t)$ and $F_y(t)$ represent the variable cutting force at time t in x-direction and y-direction and are modeled based on the regeneration effect [92][149].

Table 5. 1 Model parameters of the cutting tool in x-direction and y-direction.

	Mass(kg)	Stiffness(k_1+k_2) (N/m)	Damping ratio (%)
x	2.75×10^{-2}	8.79×10^{5}	1.39
y	2.96×10^{-2}	9.71×10^{5}	1.38

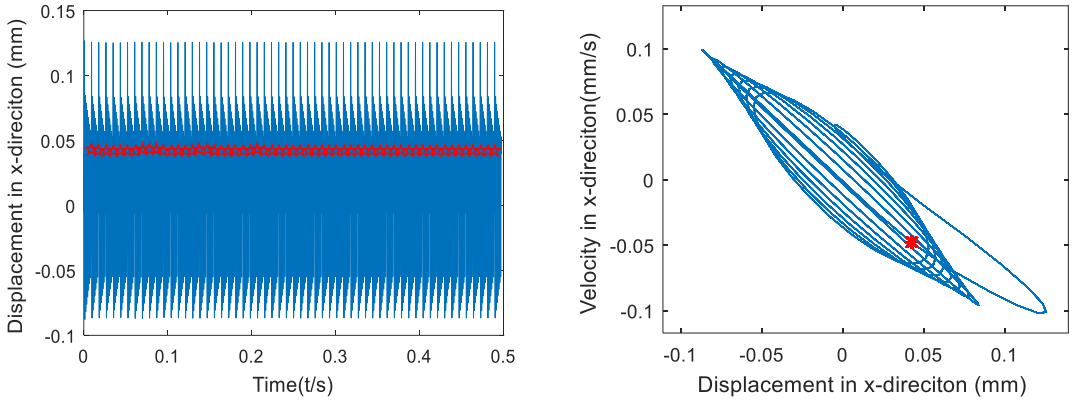
The carbide square end-cutting tool with a diameter of 12 mm, single flute, and helix angle 30° is chosen to cut the work-piece. The set machining parameters are shown in Table 5.2, respectively. The material of the work-piece is 6061-T6 aluminum.

Table 5. 2 Machining parameters.

Spindle speed (rpm)	Axial depth (mm)	Radial depth (mm)	Feed per tooth (mm/tooth)
7000	3	0.5	0.1

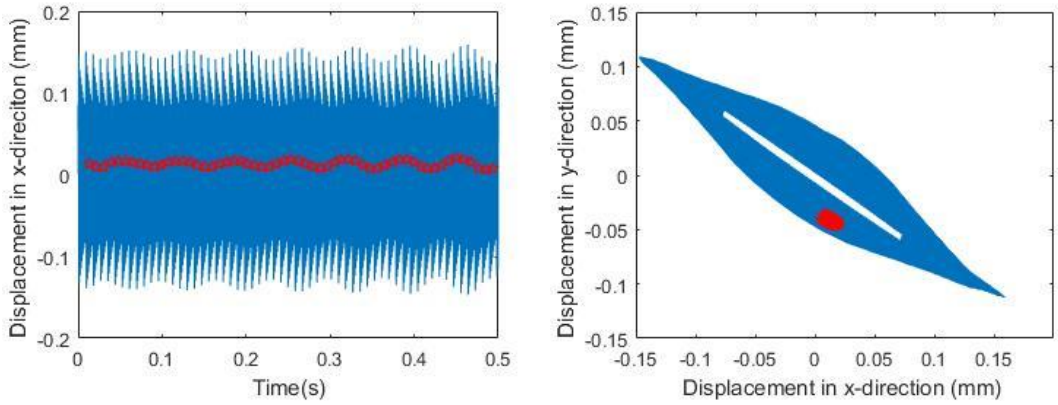
The dynamic response of the cutting tool is simulated by using the Newmark method in the time domain [150]. Figure 5.2 (a) and (b) show the displacement of the cutting tool in the x-direction and the Poincaré map without the bearing clearance fault. The red stars represent the once sampled per revolution data. It is clear from the two figures that the once sampled per revolution data is constant, which means the cutting process is stable.

During the milling process, the clearance fault may appear in the bearing of the spindle. Under the condition of the bearing clearance fault $\gamma=4\mu m$, the cutting process is simulated and plotted in Figure 5.3 (a) and (b). One can notice, the once sampled per revolution data is not constant anymore and has a slight fluctuation. A group of points appears on the Poincaré map. These phenomena reveal the appearance of the early-chatter.



(a) (b)

Figure 5. 2 Displacement response of the cutting tool at the chatter-free stage, (a): Displacement of the cutting tool in the *x*-direction, (b): Poincaré map of the cutting tool in the *x*-direction.



(a) (b)

Figure 5. 3 Displacement response of the cutting tool at the early-chatter stage, (a): displacement of the cutting tool in the *x*-direction, (b): Poincaré map of the cutting tool in the *x*-direction.

The generated chatter-free and early-chatter signals are used to evaluate the three proposed early-chatter detection methods.

5.2.2 Milling process model without machine fault

A milling process model without machine fault is also introduced and applied to generate a stable signal and period-*N* bifurcations signals. The stable signal is generated under the condition of suitable machining parameters. Whereas the period-*N* bifurcations signals are simulated under the condition of the improper machining parameters. We evaluate the performance of the period-*N* bifurcations identification method using the generated signals.

Existing research works validated the chosen milling process model without a machine fault [13][69]. Figure 5.4 shows this dynamic milling model with a 3-degree of freedom (DOF). The cutting tool is regarded as a 2-DOF mass-spring-damper system in the feed (*x*-) direction and the cross-feed (*y*-) direction. The work-piece is installed on a specific flexure with the thin-walled structure. The flexure with the work-

piece is seen as a single DOF mass-spring-damper system in the *y*-direction. The dynamic property of the flexure in the *x*-direction is ignored because the stiffness of the flexure in the *x*-direction is far higher than that in the *y*-direction.

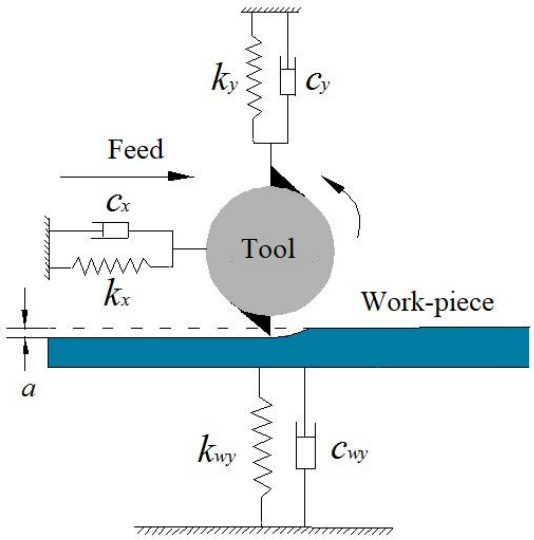


Figure 5. 4 Dynamic milling model without machine fault.

According to the machining theory and D'Alambert principle, the dynamic differential equations of the milling process system are:

$$\begin{bmatrix} m_x & & \\ & m_y & \\ & & m_{wy} \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{y}_w \end{bmatrix} + \begin{bmatrix} c_x & & \\ & c_y & \\ & & c_{wy} \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{y}_w \end{bmatrix} + \begin{bmatrix} k_x & & \\ & k_y & \\ & & k_{wy} \end{bmatrix} \begin{bmatrix} x \\ y \\ y_w \end{bmatrix} = \begin{bmatrix} F_x \\ F_y \\ F_{wy} \end{bmatrix}, \tag{5-2}$$

where m_x, m_y, c_x, c_y, k_x, and k_y represent the model mass, model damping, and model stiffness of the cutting tool in the *x*- and *y*-directions, respectively. The variables x, y, $\dot{x}$, $\dot{y}$, $\ddot{x}$, and $\ddot{y}$ represent the displacement, velocity, and acceleration of the cutting tool in the *x*- and *y*-directions, respectively. F_x and F_y represent the variable cutting force acting on the cutting tool in the *x*- and *y*-directions, respectively, which are modeled based on the regeneration effect [92][149]. The variables m_{wy}, c_{wy}, and k_{wy} represent the model mass, the model damping, and the model stiffness of the flexure with the work-piece in the *y*-direction, respectively. The variables y_w, $\dot{y}_w$, and $\ddot{y}_w$ represent the displacement, velocity, and acceleration of the work-piece in the *y*-direction, respectively. F_{wy} represents the variable cutting force acting on the work-piece in the *y*-direction, which is also known as the reacting force of F_y.

The selected cutting tool is a single-flute end-mill with a helix angle of 30 degrees and a diameter of 19.1 mm. Table 5.3 shows the natural frequency, damping, and stiffness of the cutting tool in the *x*- and *y*-directions [69].

The flexure with the work-piece is regarded as a single DOF dynamic system. Table 5.4 shows the natural frequency, damping, and stiffness of the flexure with the work-piece in the *y*-direction under three conditions. The material of the work-piece is also 6061-T6 aluminum.

We employ the milling process model without the faults mentioned above to generate the stable and period-N bifurcations signals. According to the machining parameters shown in Table 5.5, we can simulate the cutting process and obtain the stable, period-2, and period-3 bifurcation signals.

Table 5. 3 Equivalent parameters of the cutting tool head.

	Natural frequency (Hz)	Stiffness(k_1+k_2) (N/m)	Damping ratio (%)
Cutting tool in the *x*-direction	1055	4.2×10^7	4.5
Cutting tool in the *y*-direction	1055	4.2×10^7	4.5

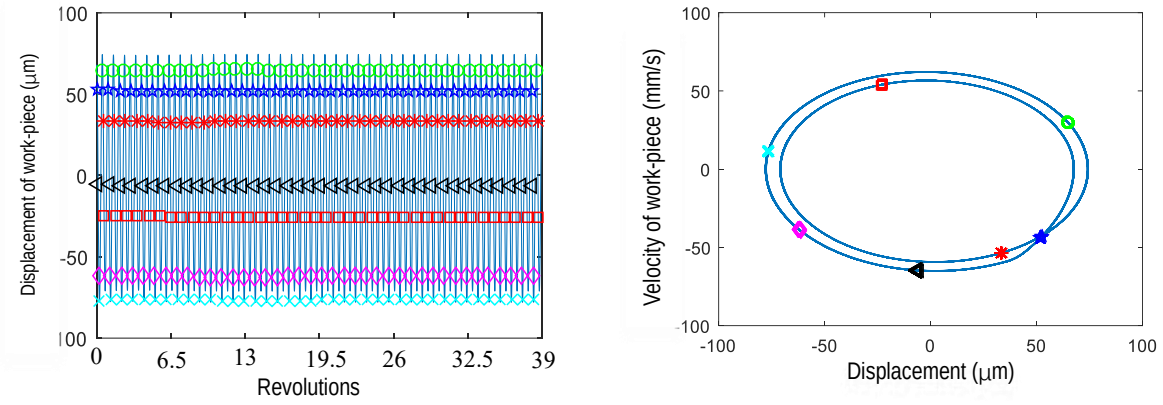
Table 5. 4 Flexure dynamics in the y-direction

	Natural frequency (Hz)	Stiffness($k1$+$k2$) (N/m)	Damping ratio (%)
Stable	130.1	2.1×106	1.47
Period-2	83	9.0×105	2.00
Period-3	163.0	5.6×106	1.08

Table 5. 5 Machining parameters.

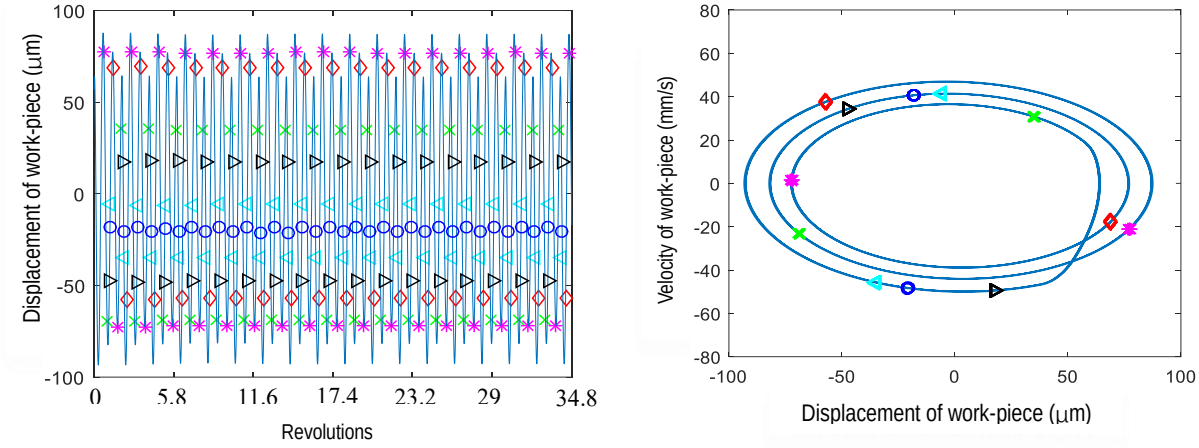
	Axial depth (mm)	Radial depth (mm)	Feed per tooth (mm/tooth)	Spindle speed (rpm)
Stable	1	2	0.1	4000
Period-2	2	1	0.1	3486
Period-3	4.5	5	0.1	3800

The setting of the simulation step size allows collecting 400 samples in one revolution. Figure 5.5, Figure 5.6, and Figure 5.7 show the displacements with MSPR data of work-piece in the y-direction and corresponding Poincaré maps. Three displacements are related to the stable, period-2, and period-3 bifurcation signals, respectively. We can see that each set of OSPR data in MSPR data is constant for the stable signal. However, the period-2 and period-3 bifurcation signals have different fluctuations in each set of OSPR data.



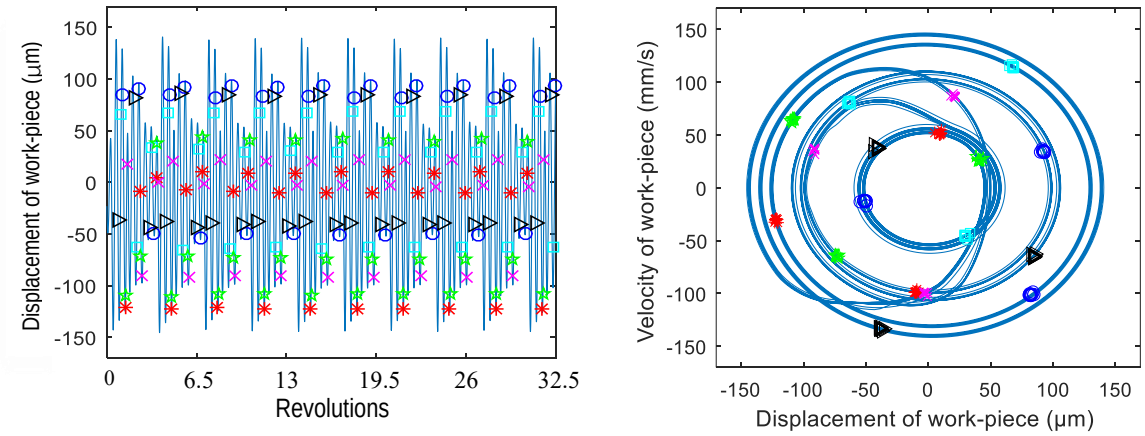
(a) (b)

Figure 5. 5 Stable displacement and corresponding Poincaré map.



(a) (b)

Figure 5. 6 Period-2 bifurcation displacement and corresponding Poincaré map.



(a) (b)

Figure 5. 7 Period-3 bifurcation displacement and corresponding Poincaré map.

Each set of OSPR data depicts partial information of bifurcation. For instance, for the period-2 bifurcation signal shown in Figure 5.6, one set of OSPR data with pink star symbols has significant fluctuation, whereas another set of OSPR data with blue cycle symbols has a small fluctuation. Therefore, it is easy to find that one set of OSPR data may fail to identify the period-2 bifurcation signal. Also, the subharmonic sampling

strategy has a limitation in the aspect of identifying the period-N bifurcations.

To simulate the noisy environment, we add white noise to three generated displacement signals with the set SNR of 10 dB. Figure 5.8, Figure 5.9, and Figure 5.10 show the stable, period-2, and period-3 bifurcation signals with noise.

The dynamic response of the two milling process models mentioned above was simulated by using the numerical simulation algorithm introduced in [150].

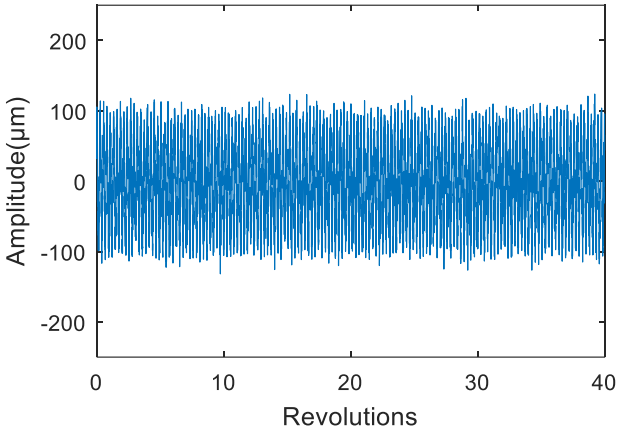


Figure 5. 8 Stable displacement with noise.

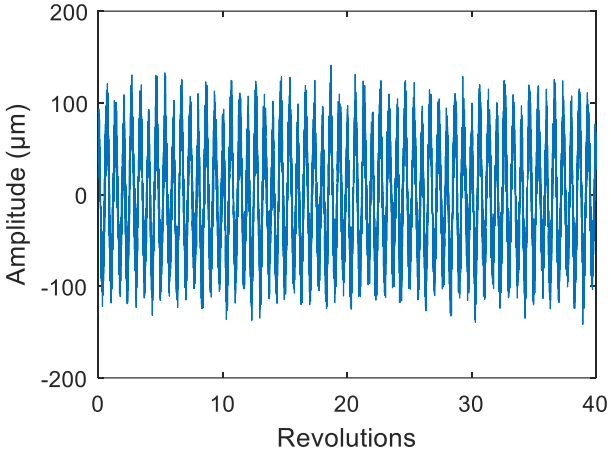


Figure 5. 9 Period-2 bifurcation displacement with noise.

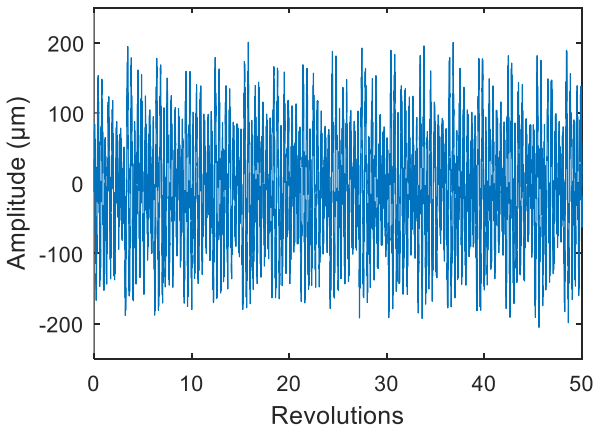


Figure 5. 10 Period-3 bifurcation displacement with noise.

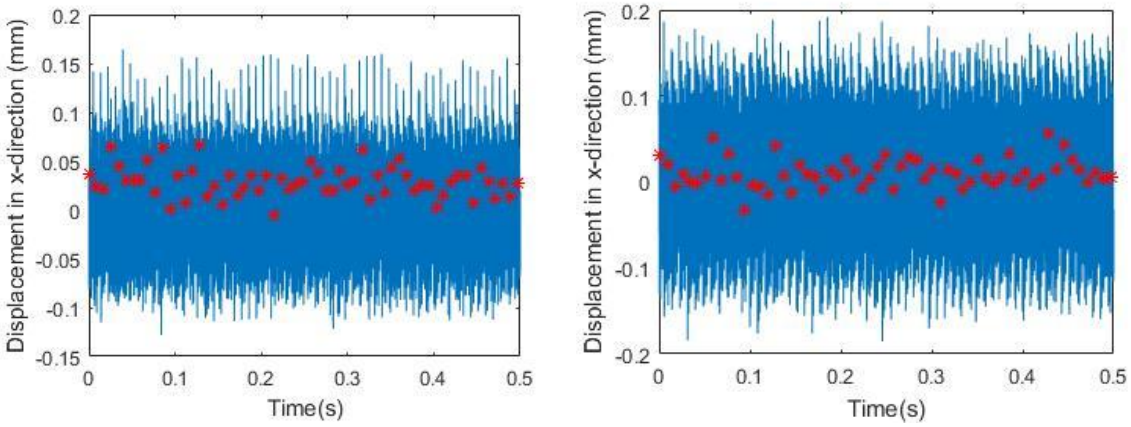
5.3 Evaluation of proposed three early-chatter detection methods

Chapter 3 introduced three early-chatter detection methods based on MaxEnt, SPRT, MPRS, M-SPRT, and SORM. In this section, these three early-chatter detection methods will be evaluated by using the milling process model previously introduced with bearing clearance fault.

5.3.1 Evaluation of the early-chatter detection method based on MaxEnt and SPRT

We used the milling process model with bearing clearance fault to evaluate the proposed early-chatter detection method based on MaxEnt and SPRT. The simulated signal is the displacement of the cutting tool in the *x*-direction. In the machining signal, the noise is inevitable. To simulate the noisy circumstance, we add the same white noise to the chatter-free and early-chatter signals. In our work, the added white noise is generated based on the early-chatter signal. The signal to noise ratio (SNR) is set to 10 dB.

According to the OSPR technique, we collected the OSPR data of the displacement in the *x*-direction using a constant angle interval of 2π synchronously. Figure 5.11 (a) and (b) depict the collected OSPR data at the chatter-free and early-chatter stages, respectively. One can see that the two series of OSPR data are close. It is difficult to detect the early-chatter using the standard deviation (SD), which was used as an indicator for chatter detection in [92][91].



(a) (b)

Figure 5. 11 Effect of noise on the cutting tool displacement at chatter-free (a) and early-chatter (b) stages.

The proposed method based on MaxEnt and SPRT is used to detect early-chatter. To implement the early-chatter detection, we need to set the prior knowledge, i.e., the time-window width, the PDF of the MaxEnt at the chatter-free and early-chatter stages, and set the two decision risk levels. Considering the early-chatter property and the detection delay, we set the time-window width to 25 revolutions for acquiring relatively sufficient chatter information. According to the experts' knowledge or a large number of experimental results, the distributions of the MaxEnt in two stages can be obtained, as shown in Figure 5.12. The distributions in both stages are closed. The mean of the MaxEnt distribution at the early-chatter stage is higher than that at

the chatter-free stage. In practice, the decision risk levels are set to 0.05 or 0.01. Thus, we set the two decision risk levels to 0.01 in early-chatter detection.

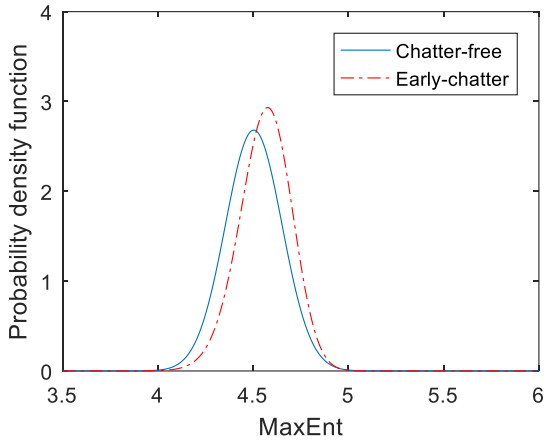


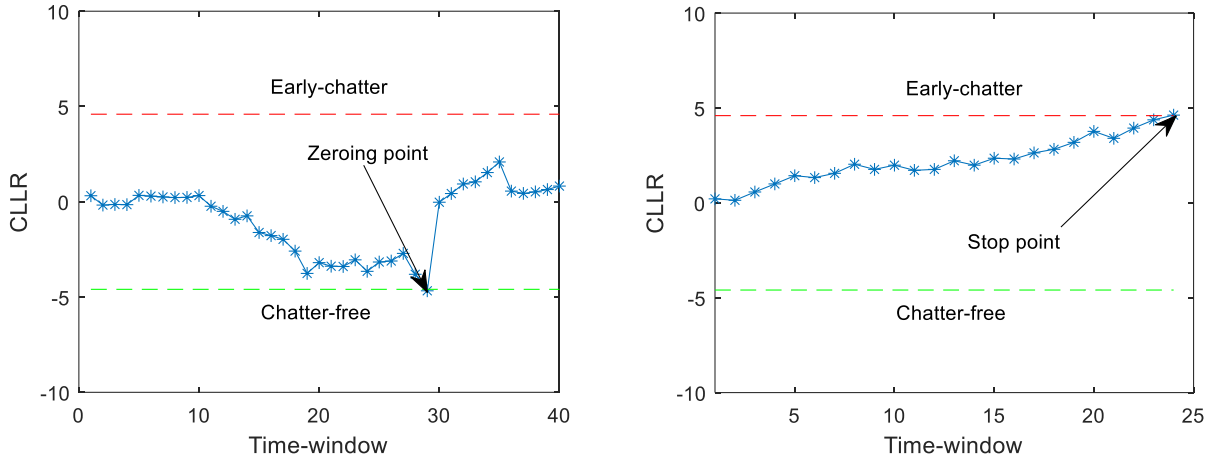
Figure 5. 12 MaxEnt PDF at the chatter-free and early-chatter stages.

The proposed method employs the MaxEnt of OSPR data as an indicator of early-chatter detection. After the OSPR of the displacement signal, the MaxEnt principle is performed on the OSPR data to extract the corresponding MaxEnt feature. The MaxEnt feature from the OSPR data is extracted using several statistics, such as mean, standard deviation, skewness, and kurtosis. Thus, the obtained MaxEnt includes sufficient chatter information relatively to only one statistic measure, such as standard deviation, and can highlight the appearance of the early-chatter more accurately.

Besides, the SPRT is used to accumulate information embedded in the MaxEnt feature for decision-making. At the early stage of the chatter, the chatter feature is unapparent due to the disturbances by the noise. Indeed, the SPRT has the property to accumulate the unapparent chatter information embedded in the MaxEnt feature, and that can help to make decisions reliably.

Figure 5.13 shows the results of early-chatter detection. The red and green horizontal dashed lines represent the upper and lower thresholds, respectively. When the cumulative log-likelihood ratio (CLLR) lies below the lower threshold, it means that the milling process is at the chatter-free stage. Meanwhile, the preceding CLLR is set to zero and continues monitoring the milling process for detecting the early-chatter. When the CLLR crosses the upper threshold, it means the early-chatter has occurred. When the CLLR is between the up and down thresholds, no decision is made, and the monitoring continues with the detection procedure of early-chatter. For the chatter-free data shown in Figure 5.13(a), the CLLR fluctuates around the lower threshold. It is the result of the resetting of CLLR to zero when the lower threshold is reached. For instance, after the lower threshold is reached at the 29^{th} monitoring result, the cutting state is recognized as the chatter-free. The cutting state monitoring continues with the detection procedure. The 29^{th} CLLR is reset to zero to remove the influence of accumulation result on the later detection. After the reset of the CLLR, the detection procedure continues to monitor the cutting state until the upper threshold is reached. For the early-chatter

data plotted in Figure 5.13 (b), the upper threshold is crossed at the 24th monitoring result. It means the early-chatter is recognized successfully.



(a) (b)

Figure 5. 13 Chatter detection using SPRT at (a) chatter-free and (b) early-chatter.

From the simulation experiment results, one can conclude the effectiveness of the proposed cumulative method for detecting early-chatter.

The OSPR data is constant at the stable stage while it has a fluctuation at the chatter stage regardless of the operating conditions. Therefore, the proposed method is independent of the operating conditions. Moreover, this method has much less computational complexity than methods based on FFT because the OSPR reduces the amount of sampled data significantly.

The proposed MaxEnt has a low computational cost for real-time early-chatter detection. In this method, the process of estimating PDF based on the MaxEnt feature takes most of the running time. Therefore, we introduced a numerical method for estimating the PDF first. Then, we analyzed the time complexity of the MaxEnt principle with the first four moments.

The estimated PDF based on the MaxEnt feature with the first four moments is represented by using the Lagrange multipliers. For a general case, it is impossible to obtain an analytical PDF. Therefore, some iterative algorithms are proposed to determine the approximate PDF, such as Newton's method, gradient descent, and improved iterative scaling. In this work, we employed Newton's method to estimate the PDF. Generally, the procedure consists of three steps as follows: expend the nonlinear equations (3-15) in Taylor's series around trial values of λ, drop the quadratic and higher-order terms, and solve the remaining linear equations iteratively [151]. The iteration process continues until a preset error is reached. According to the first four moments used in this work, 13 integrals are calculated for determining the Lagrange multipliers in each iteration process. The univariate Simpson's method is used to calculate these integrals.

It is necessary to consider the time complexity of the MaxEnt principle for real-time chatter detection implementation. Since the iteration needs to be produced by using Newton's method to satisfy the preset error and integrals need to be calculated using univariate Simpson's method, we relied on the time

complexity study of these two methods for analyzing the complexity of the MaxEnt.

Using the standard notation O in computer science for expressing the time complexity of an algorithm, the time complexity for the univariate Simpson's method is $O(q)$, where q is the number of subdivisions between up and down bounds of integration. Besides, the time complexity of determining the Lagrange multipliers using Newton's method with specified precision depends on the given initial approximation and the application.

Lastly, the time complexity of the MaxEnt principle with the first four moments is acceptable for real-time chatter detection implementation. Although the computational cost of the MaxEnt principle is high, it is used for complex models with several variables, such as in language processing, for instance [152]. For models involving a few variables, peculiarly one variable, the computational cost of the MaxEnt principle is low. The proposed method with the MaxEnt feature-based reliability model considers only one variable. Then, the computational process has low complexity. So the time complexity of the MaxEnt principle with the first four moments is acceptable for real-time chatter detection.

To highlight the low computational cost of the proposed MaxEnt, the displacement signal, with the time-window of 25 revolutions, is chosen to calculate the proposed MaxEnt, the approximate entropy (ApEn) [153], the sample entropy (SampEn)[154][155], and the fuzzy entropy [156][157]. These entropies are calculated on the desktop computer with a CPU of Intel Core i5-6600 3.3GHz. The computational costs of the abovementioned four entropies are about 0.0195s, 0.7417s, 0.7629s, and 1.0973s, respectively. The computational cost of the proposed MaxEnt is much lower than for the other three entropies.

Besides, to highlight the advantage of the proposed MaxEnt further, we compare the proposed MaxEnt with an advanced nonlinear and nonstationary signal processing method, i.e., improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) [158]. Based on the noise and parameters recommended by the ICEEMDAN inventors, we choose the white noise, the ratio of the added white noise of 0.2, and the ensemble number of 100 [159]. For the signal with the time-window of 25 revolutions, the computational cost of ICEEMDAN is about 26 s. It is much higher than the computational cost of the proposed MaxEnt. Also, to alleviate mode mixing and obtain optimal extraction results with the ICEEMDAN, we should select an appropriate sampling frequency considering the amplitude of the intermittent signal. This issue limits the application of the ICEEMDAN. For detail of this issue, see Appendix B.

Thus, the proposed MaxEnt is more suitable than the other three entropies and the ICEEMDAN-based methods for detecting early-chatter in the high-speed milling process.

5.3.2 Evaluation of the early-chatter detection method based on MPRS and SORM

This Subsection is dedicated to evaluating the effectiveness of the proposed early-chatter detection method based on MPRS and SORM. We also choose the milling simulation model with the bearing clearance fault

to generate the displacement signal for evaluating the effectiveness of the proposed method based on MPRS and SORM. The displacement signal used in the previous Subsection is chosen in the current Subsection. We collected the displacement data using MSPR and then calculated the relating standard deviation of the collected MSPR data. The calculated relating standard deviation is seen as the chatter feature.

The first four moments of the standard deviation of the MSPR data in the most recent cutting process are estimated according to equation (3-3) in chapter 3. Based on these estimated first four moments and the preset reference feature, one can derive the first four moments of the performance feature function. We used the SORM based on the MaxEnt feature to estimate the performance feature function and to calculate the FP. Figures 5.14 (a) and (b) show the estimated performance feature function at the chatter-free and early-chatter stages, respectively. Hereafter, we display the eight time-windows (segments) for each situation, with the time-window chosen equal to 25 revolutions. The vertical line in Figure 5.14(a) and (b) pass through Z=0. The probability $p(Z<0)$, i.e., the covered area of the performance feature function on the left side area of the vertical line, is the failure probability (FP). According to the estimated PDF and the FP, we can calculate the failure hazard function (FHF) to detect early-chatter.

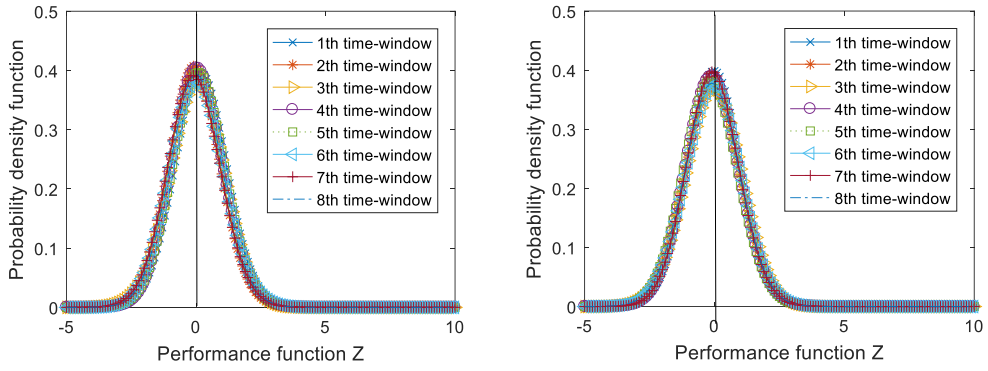


Figure 5. 14 Performance feature function at (a) chatter-free and (b) early-chatter stages.

Figure 5.15 shows the corresponding FHF at chatter-free and early-chatter stages. The blue curve with star symbols shows the FHF with fluctuations at the chatter-free stage. The fluctuation is caused by noise. The red line represents the threshold FHF_0 defined based on the 3σ principle. For the chatter-free process, the FHF is below the threshold FHF_0, which means the cutting process is under control. The pink curve with the triangle symbols depicts the FHF at the early-chatter stage. Some values of the FHF have crossed the threshold FHF_0. Thus, one can say that the early-chatter has appeared and claim that the proposed method can detect the early-chatter effectively.

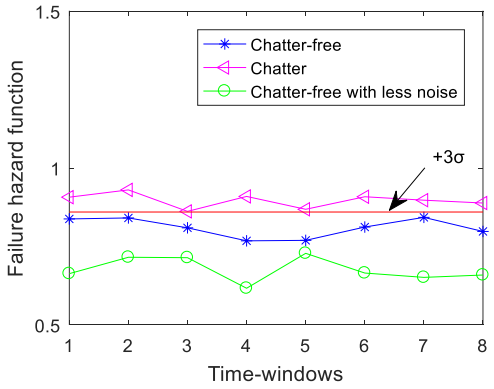


Figure 5. 15 FHF at the chatter-free and chatter stages.

Besides, we compared the proposed indicator FHF with the Kullback-Leibler (KL) divergence. The KL divergence is a tool used to compare two probability distributions. Starting with the KL divergence definition, the difference will reach zero if the two probability distributions are the same. For two different probability distributions, the KL divergence is not equal to zero. Furthermore, the KL divergence increases with the difference between the two probability distributions. The KL divergence has been applied for fault detection in dynamic systems [160]. For early-chatter detection, we can measure the difference between the estimated and the preset PDF. However, the KL divergence may not indicate the chatter appearance reliably since the KL divergence cannot distinguish the relative location of the two probability distributions. The KL divergences at the chatter-free and early-chatter stages are shown in Figure 5.16. There are two cases at the chatter-free stage. The first one is that the adding SNR is 10 dB. The second one has less noise, and its SNR is 10.5 dB. According to Figure 5.16, one can find that KL divergence fails to detect chatter. The decrease in the embedded noise leads to a false alarm. The KL divergence can measure the difference between the two probability distributions but cannot distinguish their relative position. During the utilization of a new cutting tool, the produced noise decreases first and then increases with the tool degradation. The KL divergence cannot meet this situation in detecting chatter. The proposed FHF can overcome the shortage of KL divergence. Comparing Figure 5.15 and Figure 5.16, we can find that the decrease in the embedded noise does not lead to a false alarm by using the proposed FHF.

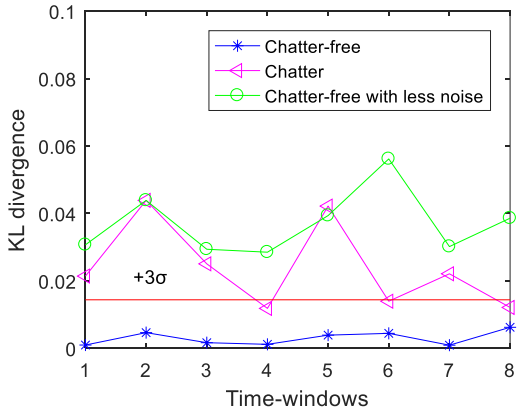


Figure 5. 16 KL divergence at the chatter-free and chatter stages.

The experimental results demonstrated the efficiency of the proposed method for the real-time detection of early-chatter. This method can detect early-chatter reliably by employing SORM and MSPR.

In this method, estimating PDF based on the MaxEnt feature also takes most of the running time as the previous method. In the previous Subsection, we analyzed the low computational cost of this method. To highlight this advantage, we assessed the chatter detection procedure on a desktop computer with Intel Core i5-6600 3.3GHz CPU. The computational cost with a time-window width of 25 is about 0.015 s. It is much less than the machine run-time 0.2143 s (the experiment with the spindle revolution speed of 7,000 rpm) in a time-window.

5.3.3 Evaluation of early-chatter detection method based on MSPR and adaptive threshold based on two decision risk levels

In this subsection, the proposed detection method of early-chatter based on MSPR and M-SPRT is evaluated using the milling process model with the bearing fault mentioned above. The proposed method includes four steps: (1) prior knowledge determination, (2) MSPR data acquisition, (3) chatter feature extraction, and (4) decision making. Therefore, we evaluate this method through the following steps:

1) Prior knowledge determination

Prior knowledge determination is essential to the detection of early-chatter. Generally, the determination of the prior knowledge is completed according to a large number of experimental studies or expert's knowledge. In this case, the sample number per revolution and the time-window width are set to 100 and 5, respectively. Both type I and type II errors are set to 0.01. Figure 5.17 shows the PDF of the MaxEnt at the chatter-free stage and early-chatter stage.

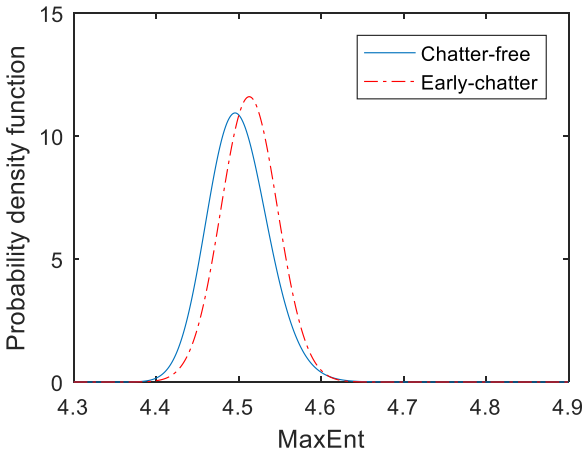


Figure 5. 17 PDF of the MaxEnt at the chatter-free stage and early-chatter stage.

2) MSPR data acquisition

For the early-chatter detection method during the high-speed milling process, the detection delay is unavoidable and should be shortened as much as possible. Using MSPR data, it is possible to reduce the

time-window width and then reduce the detection delay. In this work, MSPR data with 100 samples per revolution are collected synchronously using the MSPR technique.

3) Chatter feature extraction

In the proposed method based on OSPR and SPRT [17], the time-window width was limited within a range of 20 to 30 in order to estimate relatively accurate MaxEnt and avoid too long detection delay. In this work, multiple samples are collected in one revolution rather than one sample. Therefore, MSPR data can be employed to reduce the time-window width and to estimate the MaxEnt. We set the time-window width to 5 revolutions rather than 25 revolutions used in [17]. In this case, the detection delay is reduced significantly. MSPR data are zero-centered to obtain the data that are near to zero. Figure 5.18 (a) and (b) show the zero-centered MSPR data at the chatter-free and early-chatter stages. MaxEnt principle is used to estimate the probability distribution of zero-centered MSPR data. Based on the estimated probability distribution, the MaxEnt can be calculated as a chatter feature. Then, the M-SPRT is used to decide on whether the early-chatter has appeared.

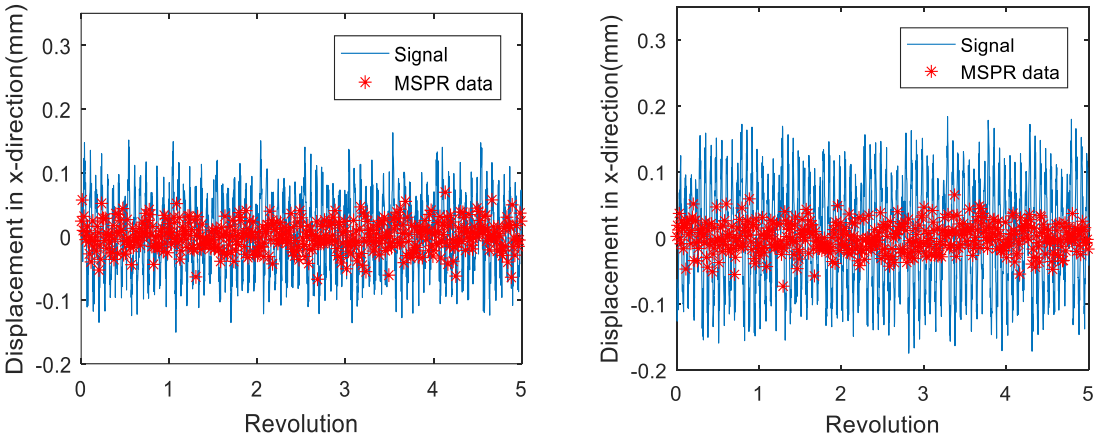


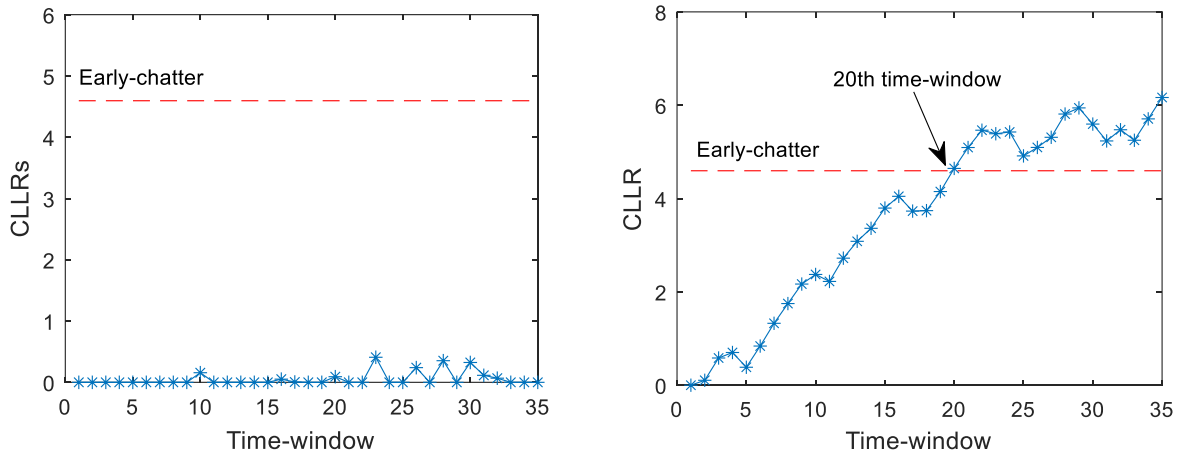
Figure 5. 18 Zero-centered MSPR data, (a) at the chatter-free stage, (b) at the early-chatter stage.

4) Decision making by using M-SPRT

According to the preset prior knowledge, the decision results of M-SPRT at the chatter-free stage and early-chatter stage are obtained, as shown in Figure 5.19 (a) and (b), respectively. All CLLRs are under the threshold under chatter-free condition, indicating the milling process is under control. Besides, some successive CLLRs that are equal to zero are observed. This phenomenon, which rarely happens in the original SPRT, occurs commonly with the M-SPRT. It is due to that the M-SPRT resets the CLLRs to zero, which are less than zero.

For the early-chatter process shown in Figure 5.19 (b), the CLLR crosses the threshold value in the 20th time-window, indicating that the early-chatter has appeared and has been detected. Due to information accumulation of the M-SPRT, the early-chatter is detected easily even though the chatter feature is not apparent. After multiple accumulations, the CLLR can be higher than the threshold value.

To illustrate the advantage of the M-SPRT in reducing detection delay, we also obtain the detection results of the original SPRT, as shown in Figure 5.20. It can be seen that the early-chatter is detected in the 22nd time-window. It is a longer detection delay than that of the M-SPRT. The reason is that the original SPRT accumulates one negative value of the LLR into the CLLR at the beginning. However, in the M-SPRT procedure, the first CLLR, which is less than zero, is reset to zero to avoid influencing the detection of the chatter-free process. Thus, without the detection of the chatter-free process, the detection delay is reduced. However, if there is no CLLR less than zero in the detection process, the detection results of the M-SPRT will be the same as that of the original SPRT. In summary, the M-SPRT can decrease the detection delay statistically due to the stochastic characteristic of CLLR.



(a) (b)

Figure 5. 19 Decision results of M-SPRT at (a) chatter-free stage and (b) early-chatter stage.

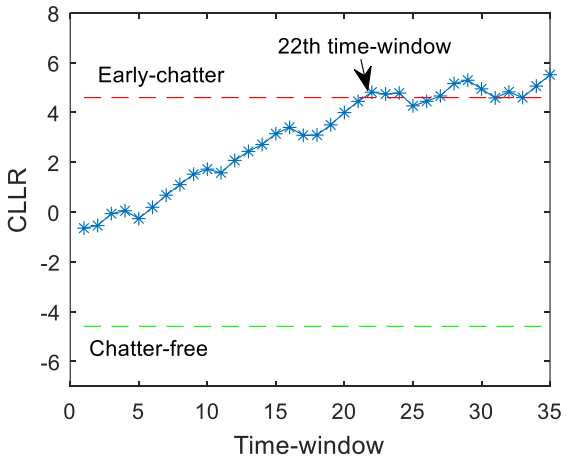


Figure 5. 20 Decision results of the original SPRT at the early-chatter stage.

In practice, the measuring instrument error and in-homogeneous material of the work-piece cause the background noise of the signal, leading to reduced detection reliability of the proposed method. To illustrate the influence of the noise environment variation on the detection reliability, we assume the SNR increases to 0.25 dB after using a high-quality measuring instrument. The decrease in the background noise leads to a decrease in the calculated MaxEnt. In this situation, a missing alarm appears. Figure 5.21 shows the

missing alarm caused by the decrease in the background noise.

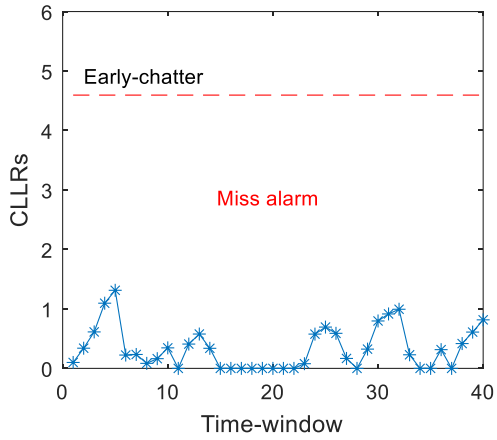


Figure 5. 21 Missing alarm caused by the decrease in the background noise.

To solve this problem, we introduce an adaptive threshold based on two decision risk levels. In this work, the prior knowledge, i.e., PDFs of the MaxEnt at the chatter-free and early-chatter stages, are reset, as shown in Figure 5.22. In comparison with the original PDFs of the MaxEnt shown in Figure 5.17, the current PDFs of the MaxEnt are shifted to the left, which is due to the decrease of the background noise. With the reset PDFs of the MaxEnt, we avoid the missing alarm, as shown in Figure 5.21. Figure 5.23 shows the detection results. The early-chatter is detected in the 5th time-window.

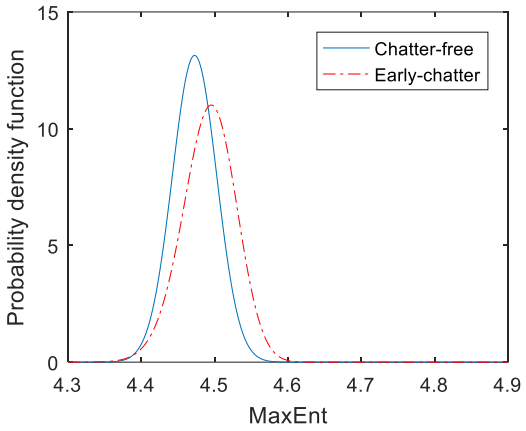


Figure 5. 22 Reset PDFs of the MaxEnt at the chatter-free and early-chatter stages.

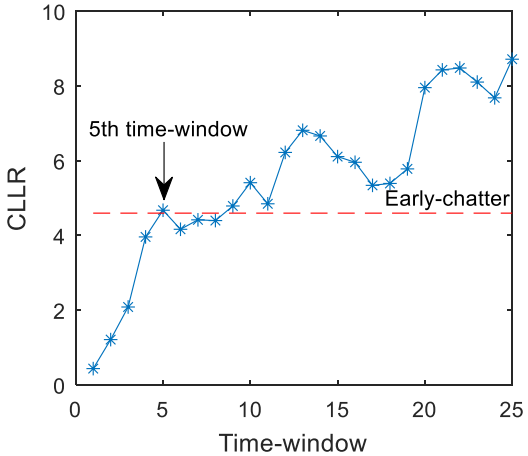


Figure 5. 23 Detection results without the missing alarm of M-SPRT.

Also, the increase in background noise may lead to a false alarm. In this case, we can reset PDFs of the MaxEnt at the chatter-free and early-chatter stages for adapting the increase of the background noise. With the reset PDFs of the MaxEnt, the false alarm can be avoided.

Similar to the previous two proposed methods, the process of estimating PDF based on the MaxEnt feature takes most of the running time as well in this method. This proposed detection method of early-chatter can realize quick early-chatter detection. To illustrate its computational cost further, we recorded the computational cost in one time-window on a desktop computer with Intel Core i5-6600 3.3GHz CPU. The computational cost was about 0.0274s, which was less than the machine run-time 0.04286 s (the spindle revolution speed was 7,000 rpm). Therefore, the proposed method can be implemented in the real-time detection of early-chatter.

By comparing the three proposed detection methods of early-chatter, we can highlight their advantages and disadvantages as follows.

(1) The early-chatter detection method based on MaxEnt and SPRT can detect early-chatter in real-time. Besides, the two decision risks are considered to increase decision accuracy. However, the detection delay in this method is too long for detecting early-chatter. The early-chatter is detected in the 24^{th} time-window. Due to 25 revolutions in one time-window, data in 600 revolutions were employed. The detection delay is the longest among the three proposed methods. Also, the OSPR data may not distinguish the chatter-free signal and the period-2 bifurcation signal. Therefore, the method based on MaxEnt and SPRT cannot detect the early-chatter related to the period-2 bifurcation reliably.

(2) The early-chatter detection method based on MPRS and SORM can detect early-chatter quickly. The experience showed that the early-chatter is detected in the first time-window. In other words, only the data of the 25 revolutions are employed. The detection delay is the shortest among the three proposed methods. Nevertheless, this method does not consider the two decision risks.

(3) The early-chatter detection method based on MSPR and M-SPRT is the most reliable among the three proposed methods. This method allows detecting early-chatter related to the period-2 bifurcation accurately and considering two decision risk levels. To face the variation of the environment noise, we can adapt the prior PDFs of the MaxEnt to avoid the missing alarm and false alarm. This method also reduces the detection delay significantly relative to the proposed method based on MaxEnt and SPRT. The early-chatter detection method based on MSPR and M-SPRT needs data in the 20 time-windows, i.e., 100 revolutions. It is less than the 600 revolutions used in the proposed method based on MaxEnt and SPRT. Therefore, we can propose to use the detection method based on MSPR and SORM for detecting early-chatter quickly while using the detection method based on MSPR and M-SPRT for detecting early-chatter reliably in the varying noise environment.

5.4 Evaluation of period-*N* bifurcations identification method

The last Section evaluates the performance of the three proposed early-chatter detection methods using the milling model with bearing clearance fault shown in Section 5.2. This Section will assess the effectiveness of the proposed Period-*N* bifurcations identification method using the milling model without a fault shown in Section 5.2.

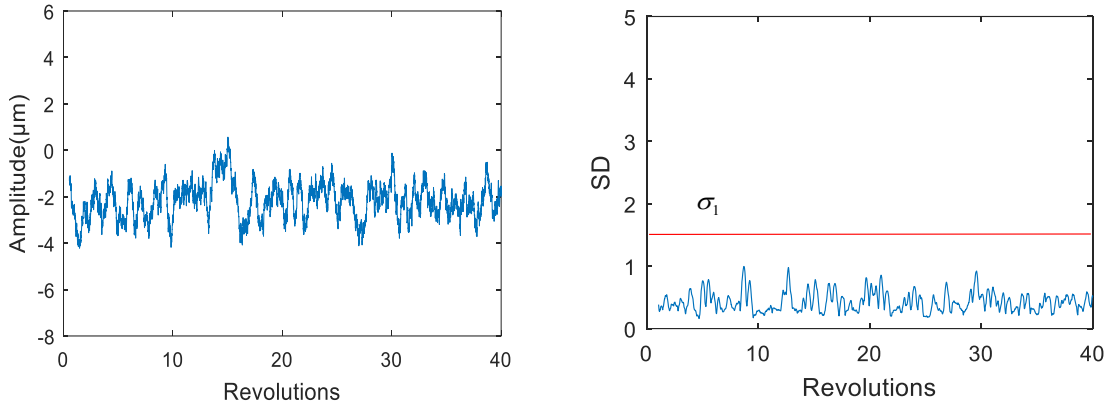
The proposed period-*N* bifurcations identification method includes two steps: period-*N* bifurcations type identification based on the simple moving average (SMA) and period-*N* bifurcations size identification based on the synchronous decomposition. Thus, we will evaluate these two steps, respectively.

5.4.1 Evaluation of period-*N* bifurcations type identification method based on the simple moving average

This Subsection evaluates the proposed period-*N* bifurcations type identification method. Before implementing the type identification of the period-*N* bifurcations based on SMA, some prior knowledge needs to be set. Here, prior knowledge includes the sample number in one revolution and a series of threshold values for the identification of various period-*N* bifurcations. In practice, the prior knowledge is set generally according to the sampling instrument and many experimental studies or expert's knowledge. We set the time-window width to *n* revolutions, where *n* is related to the type of period-*N* bifurcations. A series of threshold values are determined according to a large number of experimental studies or expert's knowledge. For the signal without the noise, the σ_N of the filtered signal can be set to zero or near zero. However, we should give the initial values to the threshold values for the signal with noise because the noise cannot be filtered thoroughly.

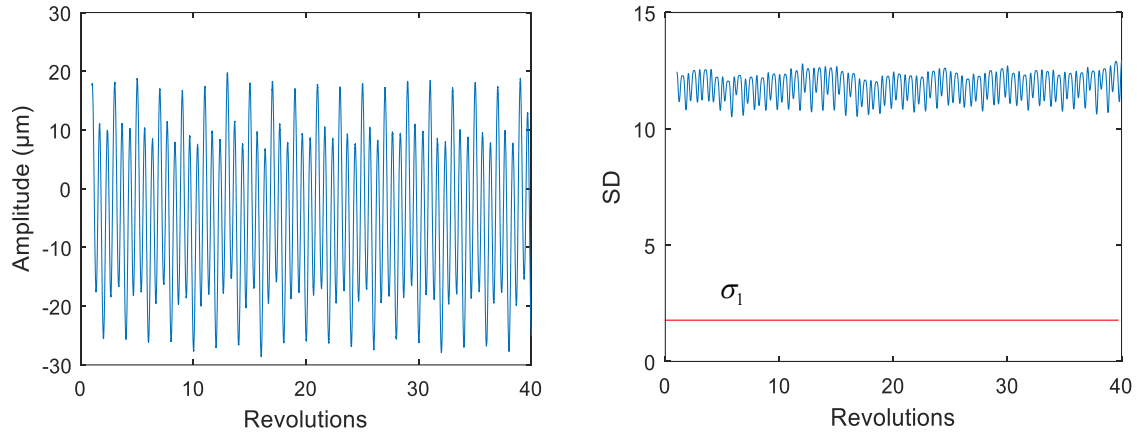
The high-order period-*N* bifurcations need a broad time-window width. The broad time-window width has a high filtering performance. Therefore, with the increase of *N*, we should set the threshold value σ_N to be small. The time-window width is an important parameter and determines the performance of the SMA. With the increase in the time-window width, the white noise is filtered more significantly. Therefore, the threshold value can be smaller. Besides, the threshold value should be significant if the signal to noise ratio (SNR) is small, vice versa. In this example, for convenience sake, we set them all to 1.5 according to our experiments. These three simulation signals with noise are used to test the type identification step of the period-N bifurcations identification method, which is based on SMA. In the beginning, we set the time-window width as one revolution for identifying the stable cutting process. For three simulation signals with noise, as shown in Figure 5.8, Figure 5.9, and Figure 5.10, the identification results are shown in Figure 5.24, Figure 5.25, and Figure 5.26, respectively. For the stable signal, the SD is less than the preset threshold value σ_1. We can see in Figure 5.24 that the cutting process is stable. For the period-2 and period-3 signals, the SD is higher than the threshold value σ_1. It means both processes are not stable. We need to identify them

continually.



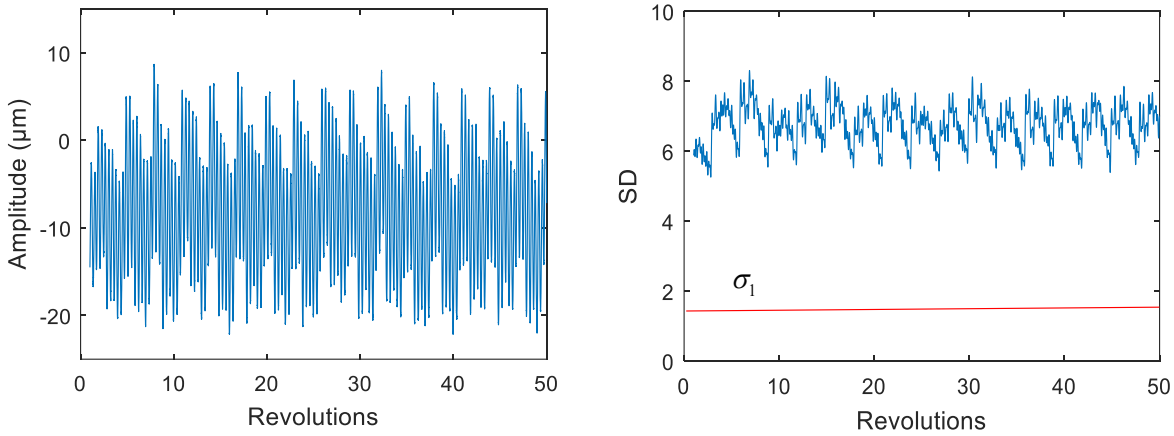
(a) (b)

Figure 5. 24 Filtered results of the SMA (a) and identification results of the stable signal (b).



(a) (b)

Figure 5. 25 Filtered results of the SMA (a) and identification results of the period-2 bifurcation signal (b).

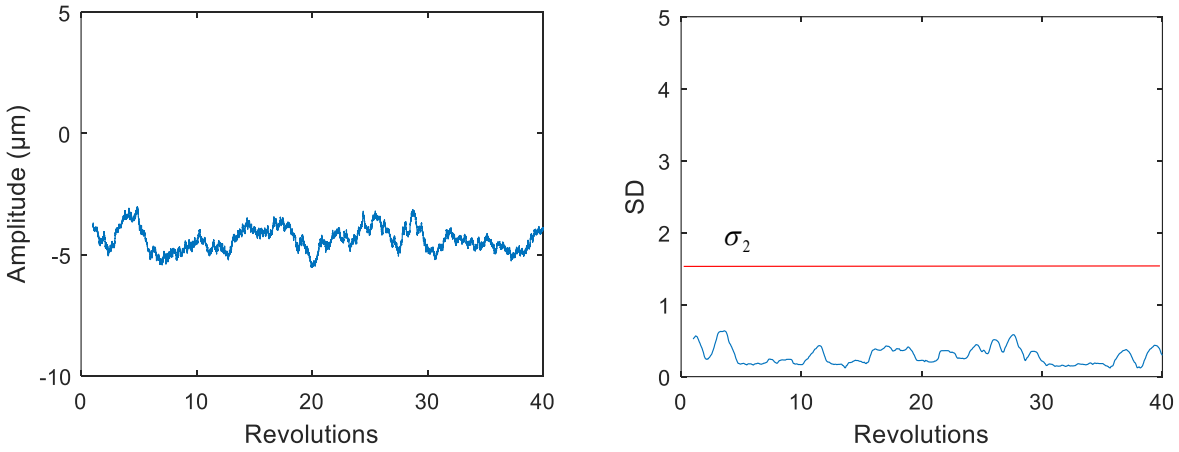


(a) (b)

Figure 5. 26 Filtered results of the SMA (a) and identification results of the period-3 bifurcation signal (b).

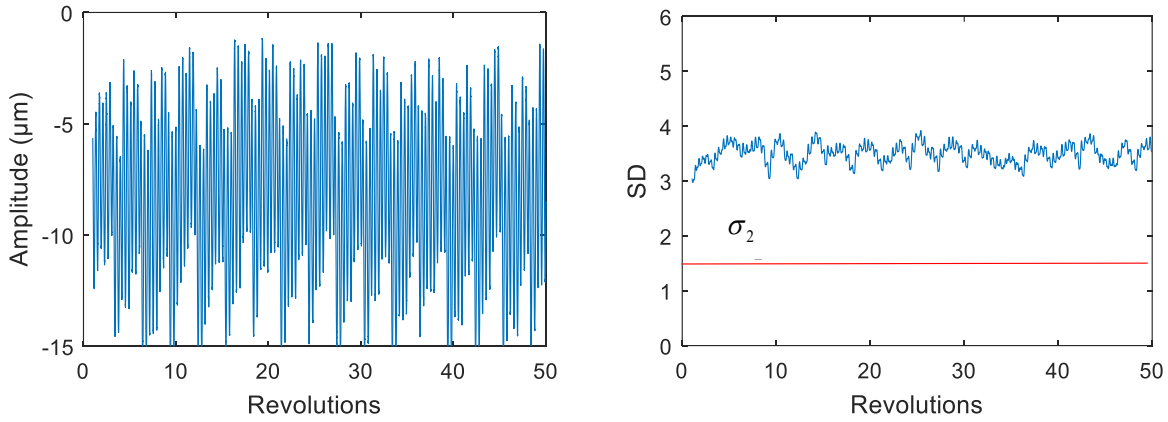
The period-2 and period-3 bifurcation signals also can be identified by using the SMA. We increase the time-window width from one revolution to two revolutions and continue to identify the remaining two signals.

Figure 5.27 and Figure 5.28 show the identification results of the period-2 bifurcation signal. The SD is less than the preset threshold value σ_2. So, we can say that the period-2 bifurcation signal has appeared. The measuring signal is the period-2 bifurcation signal. However, for the period-3 bifurcation signal, the SD is still higher than the preset threshold value σ_2. It means this signal does not belong to the period-2 bifurcation signal. The identification procedure needs to identify this signal further.



(a) (b)

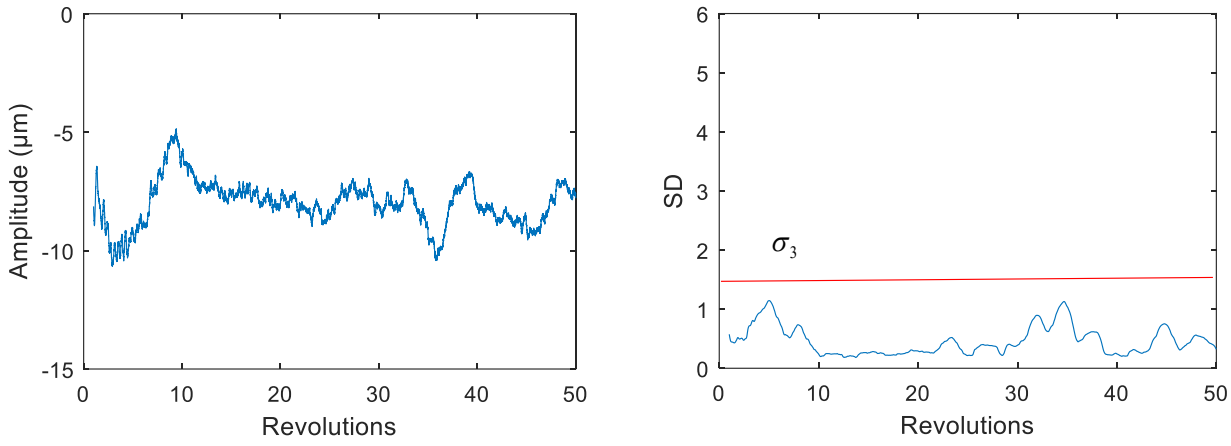
Figure 5. 27 Filtered results of the SMA (a) and identification results of the period-2 bifurcation signal (b).



(a) (b)

Figure 5.28 Filtered results of the SMA (a) and identification results of the period-3 bifurcation signal (b).

To identify the period-3 bifurcation, we increase the time-window width from two revolutions to three revolutions. Figure 5.29 illustrates the identification results of the period-3 bifurcation signal. From Figure 5.29, we can find that the SD is less than the preset threshold value σ_3. The signal can be seen as the period-3 bifurcation signal. For the period-3 bifurcation signal, the filtered results by SMA with the time-window width of 3 revolutions have fluctuation and do not look like a random signal. The reason may be that the period-3 bifurcation signal is not a steady-state signal and has a transient component.



(a) (b)

Figure 5. 28 Filtered results of the SMA (a) and identification results of the period-3 bifurcation signal (b).

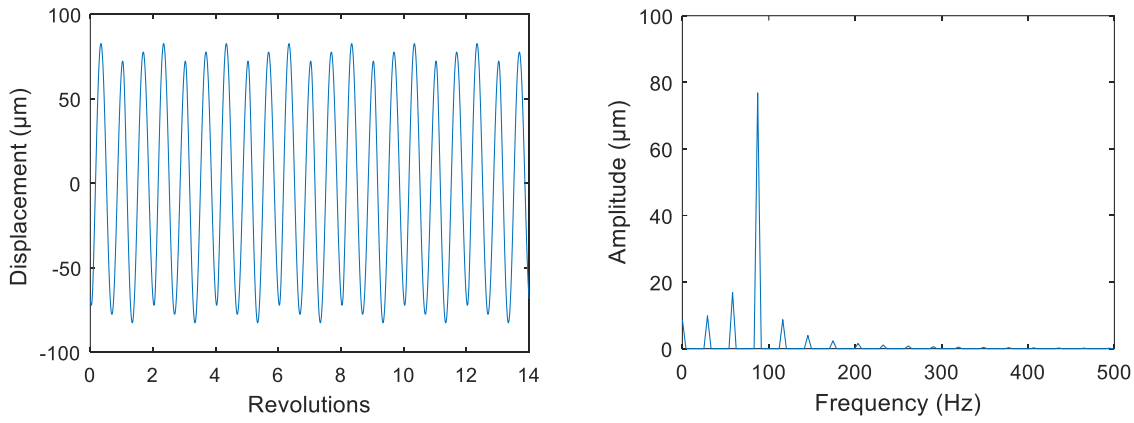
Similarly, the SMA can identify other period-N bifurcations signals. It should be noted that threshold values need to be set carefully. An improper setting of the threshold values will fail to identify the period-N bifurcations.

According to the analysis above, we conclude that the developed methodology can identify the type of period-*N* bifurcations.

5.4.2 Evaluation of period-*N* bifurcations size identification method based on synchronous decomposition

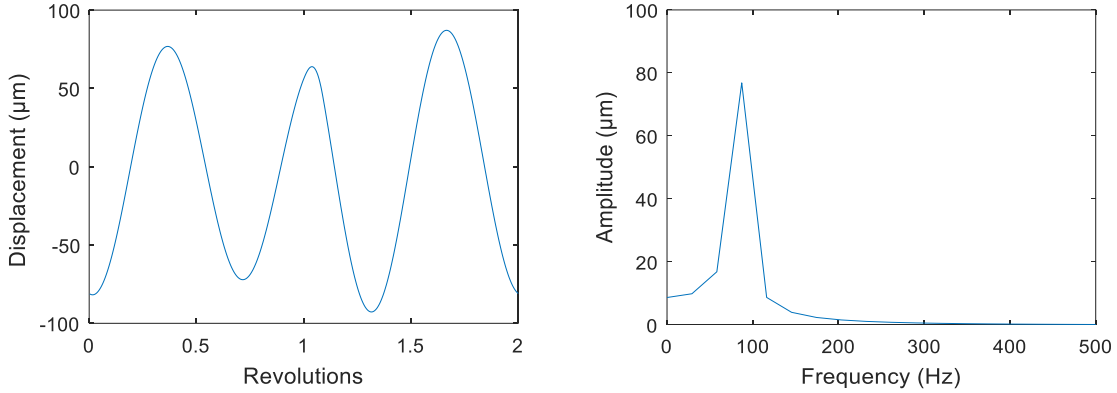
In this Subsection, the size identification part of the proposed period-*N* bifurcations identification method is evaluated using the simulation milling signals used in the previous Subsection. Besides, a period-2 bifurcation signal of the cutting tool in the *x*-direction also is employed to illustrate the advantage of the proposed method.

The period-2 bifurcation signal of the work-piece and its Fourier spectrum are shown in Figure 5.30. To show the period-2 bifurcation signal in the Fourier spectrum clearly, we set the time-window width to 14 revolutions. Figure 5.30 (b) shows that the period-2 bifurcation signal has a series of frequencies. In practice, we expect a short detection delay and a quick identification by using a short time-window. However, the resolution of the Fourier spectrum is low when the time-window is short. With a low resolution, the size of period-*N* bifurcations may not be identified accurately. For instance, the Fourier spectrum corresponding to the period-2 bifurcation signal with the time-window of two revolutions shown in Figure 5.31 (a) is illustrated in Figure 5.31 (b). We can see that the resolution of the Fourier spectrum is low. The frequencies are not shown clearly in the Fourier spectrum. In this case, it is difficult to identify the size of period-2 bifurcation using the Fourier spectrum.



(a) (b)

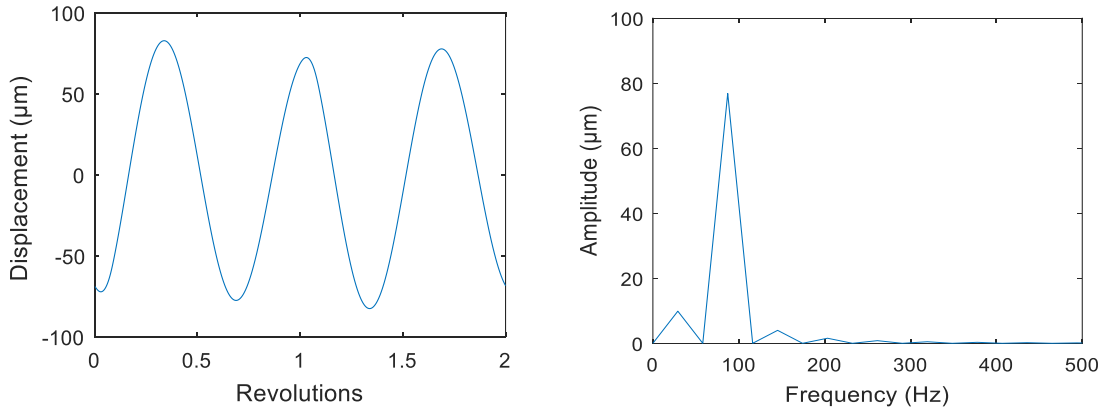
Figure 5. 29 Period-2 bifurcation signal with a time-window of 14 revolutions and corresponding Fourier spectrum.



(a) (b)

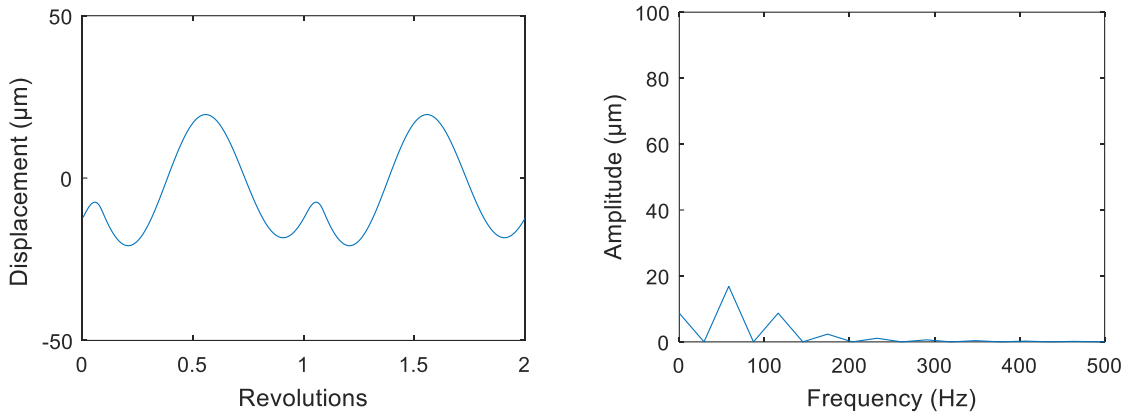
Figure 5. 30 Period-2 bifurcation signal with a time-window of 2 revolutions and corresponding Fourier spectrum.

The size identification step of the proposed period-N bifurcations identification method can accurately identify the size of period-2 bifurcation. The presented synchronous decomposition is used to decompose the period-2 bifurcation signal. Figure 5.32 and Figure 5.33 show two decomposed components and their corresponding Fourier spectrum, respectively. From these results, we find that the period-2 bifurcation signal is decomposed into the bifurcation component S_1 and the normal component S_2.



(a) (b)

Figure 5. 31 Bifurcation component S_1 with two revolutions and corresponding Fourier spectrum.

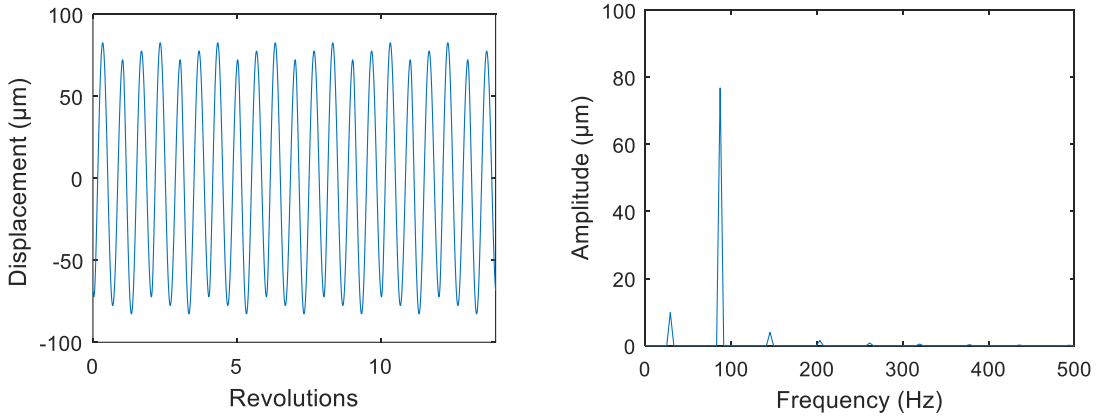


(a) (b)

Figure 5. 32 Normal component S_2 with two revolutions and the corresponding Fourier spectrum.

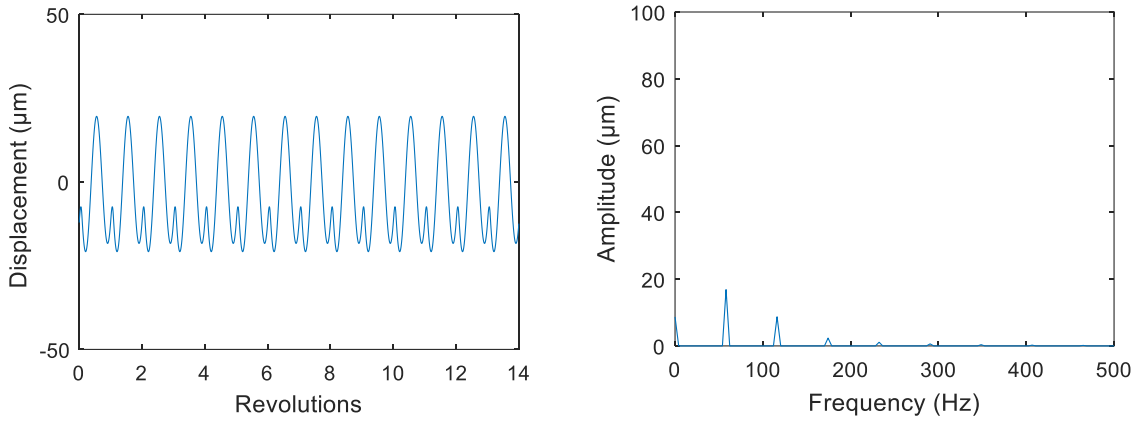
The ratio R between the SD of the bifurcation component S_1 and the period-N bifurcation signal is calculated to depict the period-N bifurcation size. The size of the period-N bifurcation is small when the R is close to zero. Also, the size of the period-N bifurcations is significant when R is close to one. For the period-2 bifurcation signal of the work-piece, the ratio R is 0.97. Thus, we can say that the period-2 bifurcation size is significant.

To show that the bifurcation component is extracted accurately, we increase the time-window width to obtain a high resolution for the Fourier spectrum. The time-window width is increased from two revolutions to 14 revolutions. Figure 5.34 and Figure 5.35 show two decomposed signals and their Fourier spectrum.



(a) (b)

Figure 5.33 Bifurcation component S_1 with 14 revolutions and corresponding Fourier spectrum.



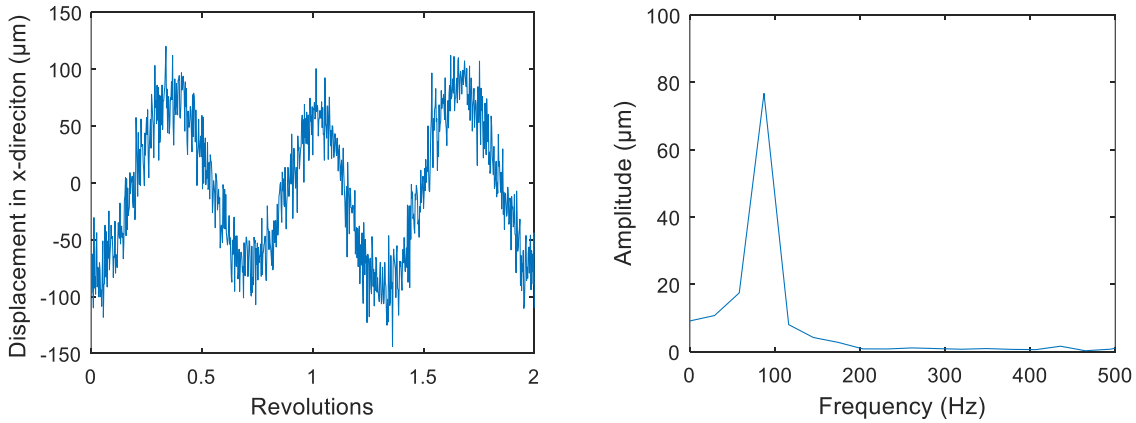
(a) (b)

Figure 5.34 Normal component S_2 with 14 revolutions and corresponding Fourier spectrum.

Figure 5.34 and Figure 5.35 show that the period-2 bifurcation signal of the work-piece is decomposed into the bifurcation component S_1 and the normal component S_2 completely.

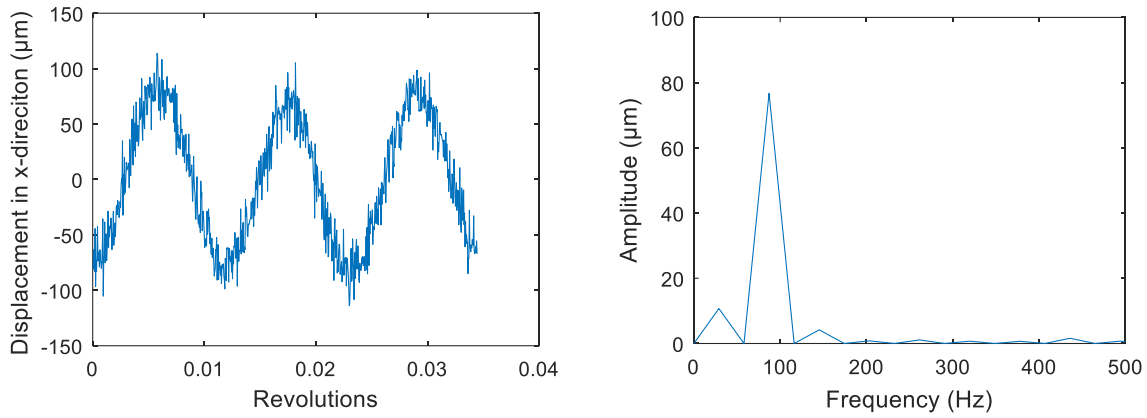
In practice, the noise is unavoidable in the measuring signal. To simulate the noisy environment, we add white noise to the generated cutting signal with the set SNR of 10 dB. Figure 5.36 displays the period-2 bifurcation signal with the noise and corresponding Fourier spectrum. The size of the period-2 bifurcation cannot be identified from the time domain and frequency domain.

The synchronous decomposition is used to decompose the period-2 bifurcation signal with the noise. Figure 5.37 and Figure 5.38 show the decomposed signals and their Fourier spectrum. The period-2 bifurcation signal with noise is decomposed into the bifurcation component S_1 and the normal component S_2. The ratio R between the SD of the bifurcation component S_1 and the period-2 bifurcation signal is calculated to depict the size of period-2 bifurcation. The size of the bifurcation is identified as 0.94.



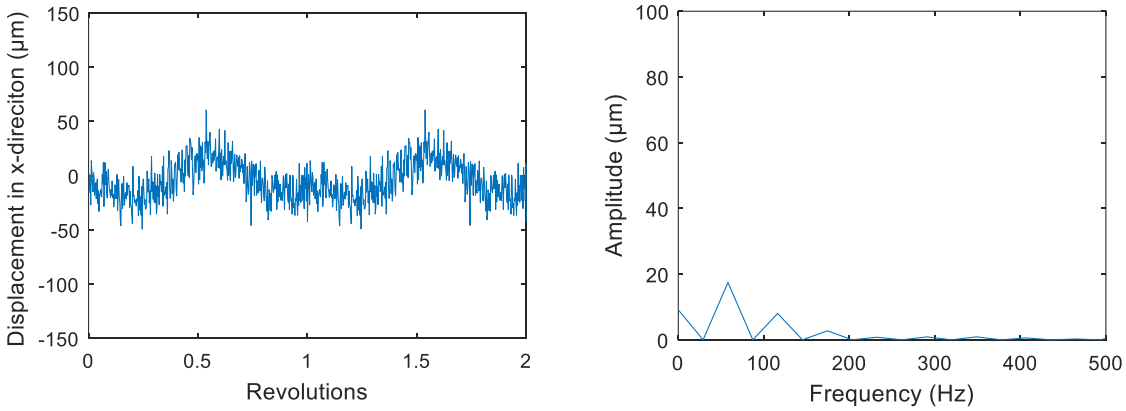
(a) (b)

Figure 5.35 Period-2 bifurcation signal with noise and corresponding Fourier spectrum.



(a) (b)

Figure 5.36 Normal component S_1 with noise and corresponding Fourier spectrum.



(a) (b)

Figure 5. 37 Bifurcation component S_2 with noise and corresponding Fourier spectrum.

In a noisy environment, the size of the bifurcation is decreased by 3%. The reason is that the added noise in the period-2 bifurcation signal also is separated into two signals S1 and S2. Figure 5.39 and 5.40 show the added noise and two separated noises, respectively. The ratio R_1 between SD of the separated noise in the bifurcation component S_1 and the added noise is about 0.71. The ratio R_2 between SD of the separated noise in the normal component S_2 and the added noise is about 0.73. The intensity of the two decomposed noises is almost the same. The added noise reduces the size of the period-2 bifurcation because the ratio R between the SD of the bifurcation component S_1 and the period-2 bifurcation signal is more significant than 0.71. On the opposite, the added noise will amplify the size of the period-2 bifurcation if the ratio R between the SD of the bifurcation component S_1 and the period-2 bifurcation signal is less than 0.71.

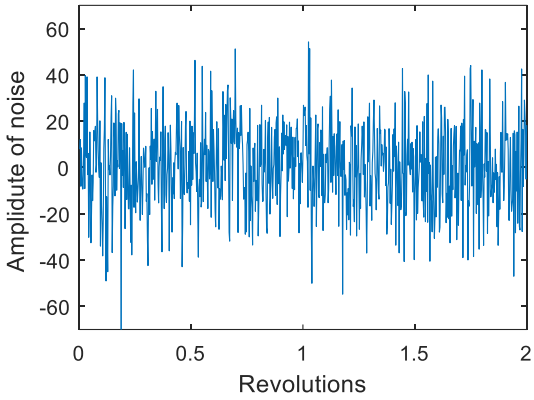
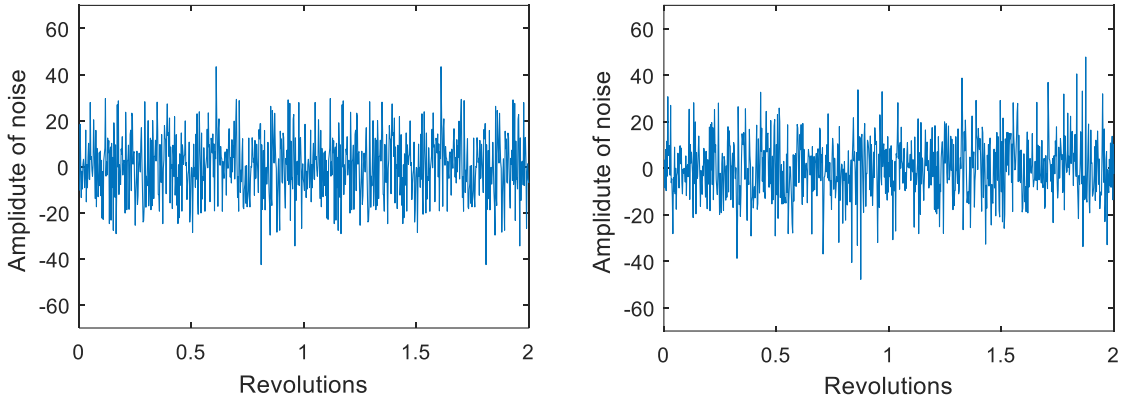


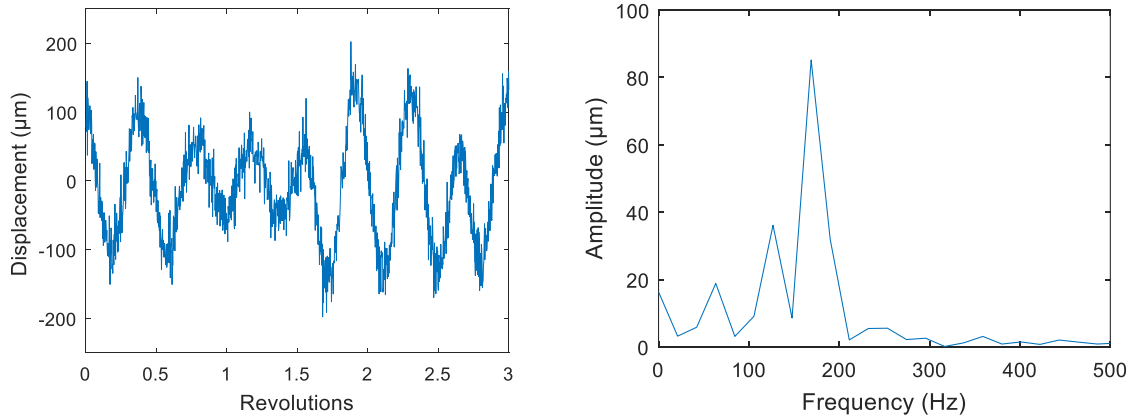
Figure 5. 38 Added noise.



(a) (b)

Figure 5. 39 Decomposed noises (a) in bifurcation component S_1; (b) in normal component S_2.

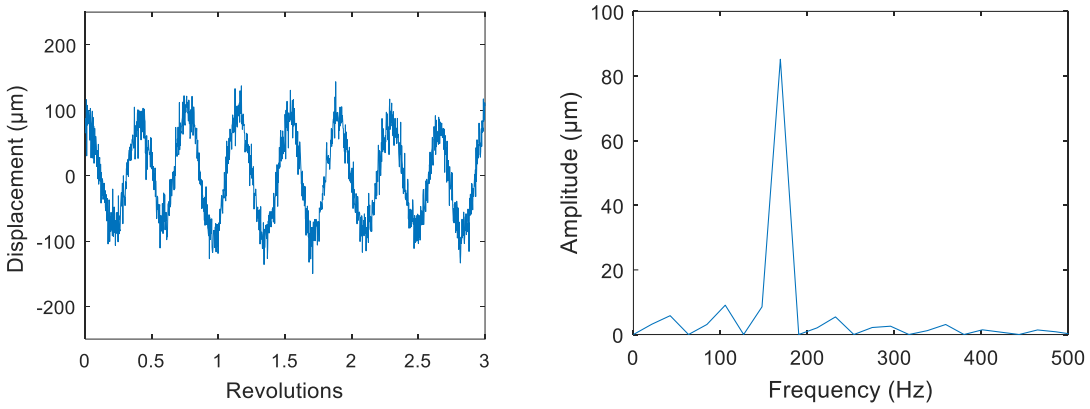
The period-3 bifurcation signal is also used to validate the size identification part of the proposed period-N bifurcations size identification method. To simulate the noise in the period-3 bifurcation signal, we also add white noise to the generated cutting signal with the set SNR of 10 dB. The period-3 bifurcation signal with added noise also is noted as S. The period-3 bifurcation signal S and its Fourier spectrum are shown in Figure 5.41. It is difficult to identify the size of the period-3 bifurcation according to the information of two domains.



(a) (b)

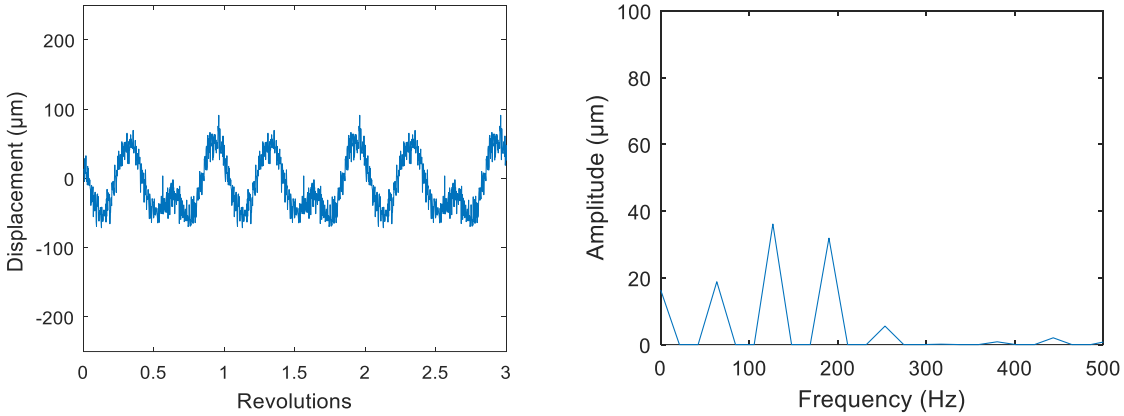
Figure 5. 40 Period-3 bifurcation signal with noise and corresponding Fourier spectrum.

The presented synchronous decomposition is used to decompose the period-3 bifurcation signal S into the bifurcation component S_1 and the normal component S_2. Figure 5.42 and 5.43 show the two decomposed components and their corresponding Fourier spectrum. We can find that the bifurcation component S_1 has been extracted from the period-3 bifurcation signal S. The ratio R between SD of the bifurcation signal S_1 and the period-3 bifurcation signal S is calculated and is about 0.7010. It can be seen that the size of the period-3 bifurcation is about 0.7010.



(a) (b)

Figure 5. 41 Bifurcation component S_1 with noise and corresponding Fourier spectrum.



(a) (b)

Figure 5. 42 Normal components S_2 with noise and corresponding Fourier spectrum.

For the period-3 bifurcation signal, the added noise also is separated into two decomposed components. Two decomposed components do not share the total noise evenly. Figure 5.44 and 5.45 illustrate the added noise and two decomposed noises in the bifurcation component and the normal component. The ratio R_1 of the SD between the noise in the bifurcation component and the added noise is about 0.5740. However, the ratio R_2 of the SD between the noise in the normal component and the added noise is about 0.8188. The less noise is decomposed into the bifurcation component. The reason is that the mean of 3 samples is used to calculate the bifurcation component S_1 in the decomposition process.

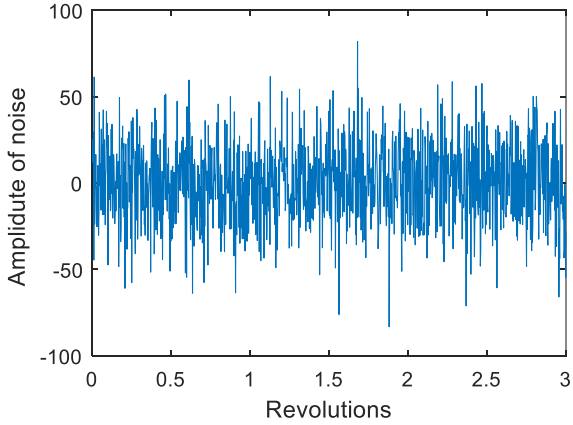
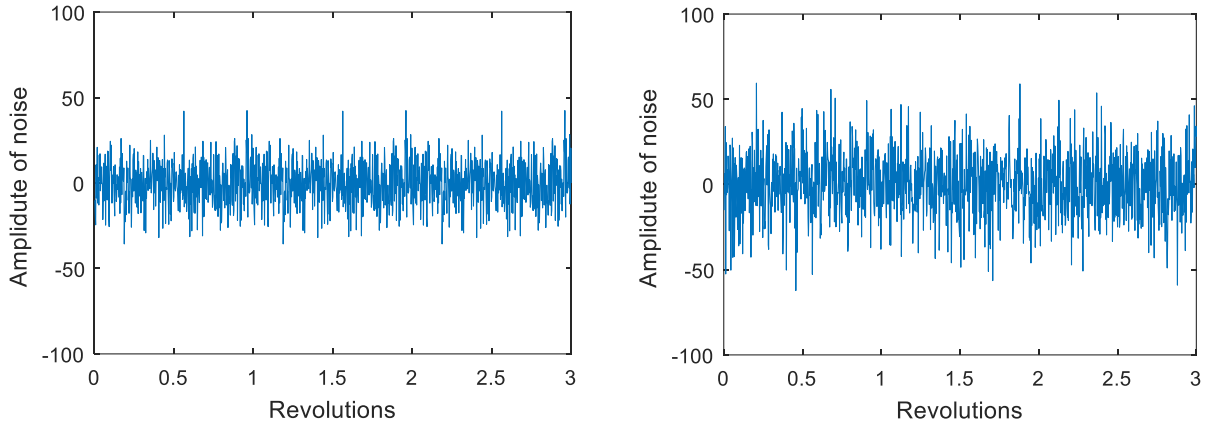


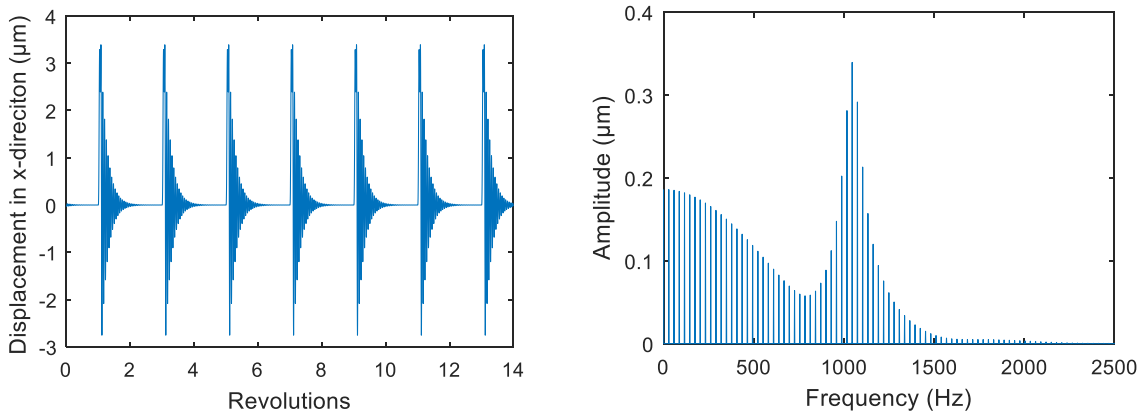
Figure 5. 43 Added noise.



(a) (b)

Figure 5. 44 Two decomposed noises, (a) in the bifurcation component S_1; (b) in the normal component S_2.

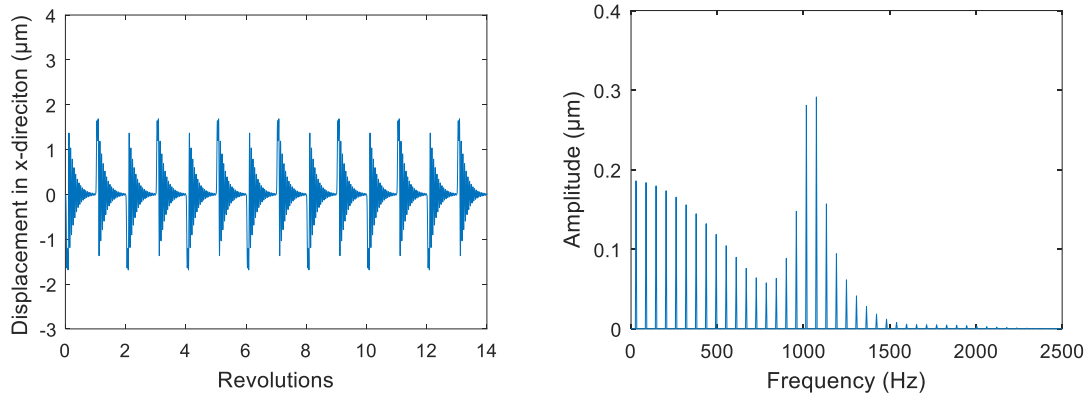
To highlight the advantage of the proposed period-*N* bifurcations identification method, we also employ the displacement signal of the cutting tool. Figure 5.46 illustrates the displacement signal of the cutting tool and its Fourier spectrum. It is difficult to identify the period-*N* bifurcations size, even if a long time-window is used, due to many frequencies in the Fourier spectrum.



(a) (b)

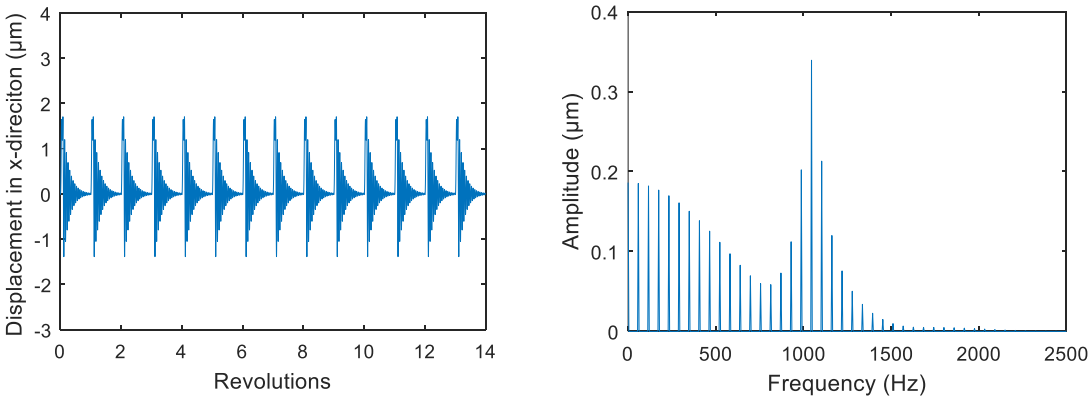
Figure 5.45 Period-2 bifurcation signal of the cutting tool and corresponding Fourier spectrum.

The proposed synchronous decomposition method can decompose the displacement signal S of the cutting tool into the bifurcation component S_1 and the normal component S_2. Figure 5.47 and 5.48 show the decomposed bifurcation component, the normal component, and their Fourier spectrum, respectively. It should be noted that the time-window width is 14 revolutions for obtaining a small resolution of the Fourier spectrum. The bifurcation component S_1 and the normal component S_2 are obtained accurately. In this situation, the ratio R between SD of the bifurcation component S_1 and the period-2 bifurcation signal S can be calculated and is about 0.7037. Therefore, the size of the period-N bifurcations is seen as 0.7037.



(a) (b)

Figure 5. 46 Bifurcation component S_1 and corresponding Fourier spectrum.



(a) (b)

Figure 5. 47 Normal component S_2 and corresponding Fourier spectrum.

From the above, it can be seen that the proposed period-N bifurcations identification method can identify the type and size of the period-N bifurcations effectively.

5.5 Conclusions

The three proposed early-chatter detection methods and the period-N bifurcations identification method were validated by using two milling models. Among the three proposed early-chatter detection methods, the early-chatter detection method based on MSPR and SORM is the fastest, while the early-chatter detection method based on MSPR and M-SPRT is the most reliable. The proposed period-N bifurcations

identification method can identify the type and size of the period-N bifurcations effectively with a short time-window.

Chapter 6 Conclusions and prospects

6.1 Conclusions

6.2 Prospects

6.1 Conclusions

This thesis has addressed the issues of real-time early-chatter detection and the period-*N* bifurcations identification based on the cumulative approach of diagnosis. Aimed at the difficulties of early-chatter detection and period-*N* identification, we proposed three real-time early-chatter detection methods and an accurate period-*N* bifurcations identification method. The main results presented in this thesis can be concluded as follows:

Early-chatter detection

(1) An early-chatter detection method based on MaxEnt and SPRT.

At the early-chatter stage, the chatter feature is inapparent and submerged by the noise. It is difficult to detect early-chatter in real-time. Aimed at this problem, we proposed an early-chatter detection method based on OSPR and MaxEnt. The OSPR is used to collect OSPR data. The OSPR reduces the amount of the data significantly compared to the traditional sampling approaches. Specifically, one sample is acquired during one spindle revolution synchronously. Based on the collected OSPR data, we propose to use the MaxEnt to characterize the early-chatter. Comparing with the commonly used standard deviation, the MaxEnt can incorporate several moments, such as mean, standard deviation, skewness, and kurtosis. Therefore, the MaxEnt seems more accurate than the standard deviation in the aspect of characterizing the early-chatter. To make a reliable decision, we used the SPRT to accumulate the inapparent chatter feature. The early-chatter can be detected through accumulating information on the inapparent chatter feature. Also, the threshold based on two decision risk levels is used to obtain more reliable detecting results.

The experiments validated that the early-chatter detection method based on OSPR and MaxEnt can detect early-chatter with detection performance at the 24^{th} time-window. In other words, the data in 600 spindle revolutions is used for early-chatter detection. The proposed method has a much low computational cost for early-chatter detection, comparing with other existing methods based on features such as approximate entropy (ApEn), sample entropy (SampEn), fuzzy entropy, and ICEEMDAN.

The early-chatter detection method based on OSPR and MaxEnt uses fewer data in early-chatter detection. Therefore, this method can be used in remote early-chatter detection with wireless sensors.

(2) Quick early-chatter detection method based on MSPR and SORM.

Although the method based on OSPR and MaxEnt can detect early-chatter in real-time, it may fail to detect the early-chatter associated with the period-2 bifurcation. The OSPR may fail to distinguish between the chatter-free and the chatter related to the period-2 bifurcation signals. Also, this method is limited for some

applications due to its long detection delay. A short detection delay is desired for real-time early-chatter detection.

To deal with these two problems, we proposed the second method for quick early-chatter detection based on MSPR and SORM. The MSPR is used to acquire multiple sets of OSPR data. Then, the bifurcation information can be acquired sufficiently. The second-order reliability method (SORM) is then proposed to derive the failure hazard function (FHF) for detecting the appearance of the early-chatter. In reliability analysis, it is a challenge to focus on the design aspect (or the most probable failure point). To avoid focusing on design consideration, we proposed to conduct the SORM based on the maximum entropy (MaxEnt) feature. The experiments have allowed evaluating the effectiveness of the proposed method. Also, this method can detect early-chatter quickly. In a case study, the data from 25 revolutions were sufficient, while the early-chatter detection method based on OSPR and MaxEnt needs the data in the 600 revolutions.

(3) Reliable early-chatter detection method based on MSPR and adaptive threshold.

This is the third method proposed for the early-chatter detection based on MSPR and adaptive threshold based on two decision risk levels. This method can detect early-chatter reliably. The main contributions can be stated in three aspects. 1) MSPR technique is proposed to acquire data. The acquired MSPR data is relatively sufficient to depict the chatter information versus the OSPR data. Also, we proposed zero-centered MSPR data to reduce the time-window width, and then the detection delay is shortened; 2) An M-SPRT is proposed to reduce the detection delay. The M-SPRT can detect early-chatter without the requirement of detecting the chatter-free process. In this situation, the detection delay is reduced statistically; 3) The adaptive threshold based on two decision risk levels is introduced to enhance early-chatter detection reliability. The adaptive threshold based on two decision risk levels considers two types of errors and avoids the assessment risk of the traditional threshold based on the small probabilities principle. Also, the adaptive threshold based on two decision risk levels adapts to the variation of the environmental noise caused by some uncontrollable influences, resulting in enhanced reliability of the proposed method. A case study has allowed assessing the proposed method's effectiveness. This method reduces the detection delay significantly, relative to the proposed method based on MaxEnt and SPRT. Compared to the proposed method based on MSPR and SORM, this method considers two decision risk levels and the background noise variation, then avoids the missing alarm and the false alarm caused by the background noise.

Among the three proposed early-chatter detection methods, the one based on OSPR and SPRT uses the smallest amount of data for early-chatter detection. Therefore, this method can be used for remote early-chatter detection with wireless sensors. The early-chatter detection method based on MSPR and SORM is the fastest, while the early-chatter detection method based on MSPR and M-SPRT is the most reliable. Therefore, we propose using the method based on MSPR and SORM to detect early-chatter quickly and using the method based on MSPR and M-SPRT to detect early-chatter reliably. Figure 6.1 illustrates the applications of three proposed early-chatter detection methods.

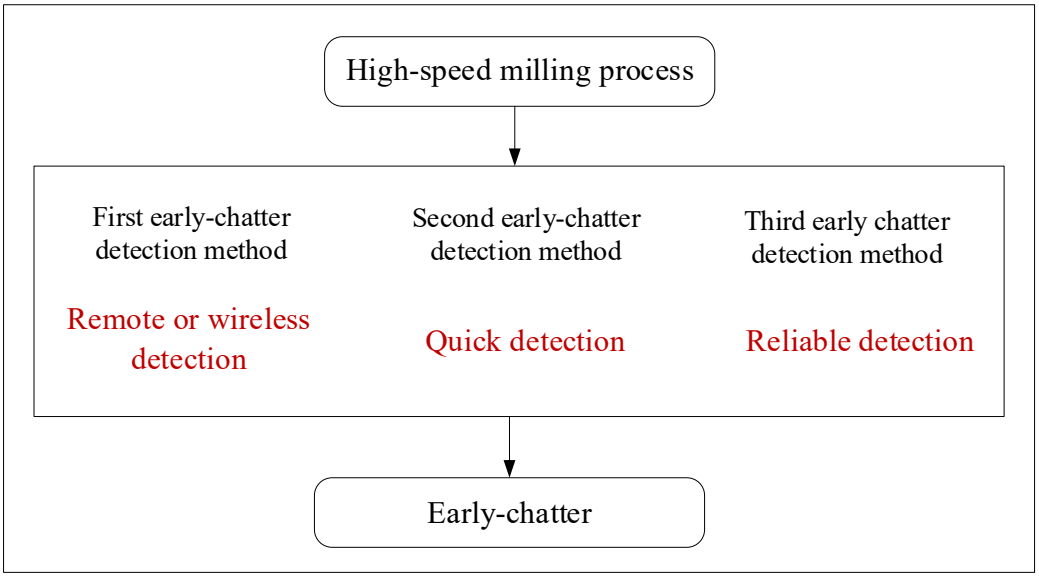


Figure 6. 1 Applications of three proposed early-chatter detection methods.

The proposed early-chatter detection methods can be applied to other dynamic systems, such as bearing, pulse energy conversion systems, and power systems. The chatter relates to an unstable phenomenon (bifurcations or chaos processes). The unstable phenomenon exists in many nonlinear dynamic systems, such as rotor-bearing systems and pulses energy conversion systems. In the rotor-bearing system, component degradation may lead to the appearance of unstable phenomena [161]. The early degradation detection of components is an essential task for the fault diagnosis and prognosis [9]. The two proposed methods based on MSPR are suitable for monitoring the incipient degradation of a component by detecting the rise of instability. According to the literature [35][36], the boundary between the stable and unstable area is a region instead of a line. In this region, the bifurcation phenomenon may have intermittent property, i.e., alternation appearance of stable and unstable phenomena. In this case, the two proposed methods based on MSPR are also suitable for their rapidity and reliability property.

Period-N bifurcations identification

The proposed period-*N* bifurcations identification method can identify the type and the size of period-*N* bifurcations. To identify the type of period-*N* bifurcations, we proposed to use the SMA. Comparing with the sub-harmonic sampling proposed in [87], the SMA can identify the type of period-*N* bifurcations reliably. To identify the size of the period-*N* bifurcations, we suggested a synchronous decomposition method. A case study and mathematical proof have allowed validating the effectiveness of the method of synchronous decomposition. The period-*N* bifurcations size can be identified using the synchronous decomposition method. Compared with the FFT, our method identifies the period-*N* bifurcations size with a narrow width of time-window. Therefore, the identification delay in our method is shorter. Also, aimed at the signal, including multiple sub-harmonic components and inter-harmonic components, it is difficult to identify the size of period-*N* bifurcations using FFT. Nevertheless, our method can identify the size of the period-*N* bifurcations accurately.

The proposed period-*N* bifurcations identification method can be used to improve productivity by identifying the period-*N* bifurcations. According to [13][90], the machining quality obtained under conditions of period-*N* bifurcation is better than that obtained under conditions of Hopf bifurcation. Thus, period-*N* bifurcations identification in machining conditions should help to improve the milling process productivity. Thus, one can improve the milling process's productivity using the proposed period-*N* bifurcations identification method. Figure 6.2 depicts the milling process scheme under the period-*N* bifurcation condition.

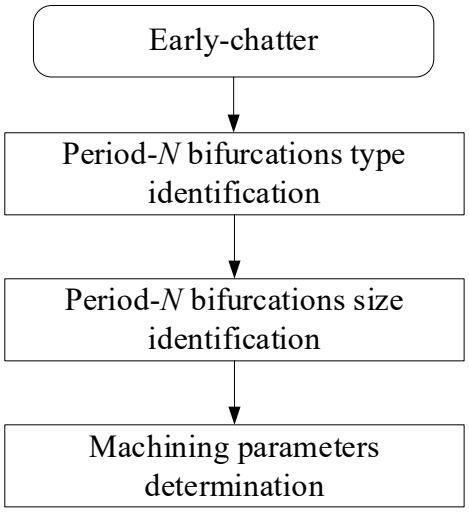


Figure 6. 2 Milling process scheme under the period-*N* bifurcations condition.

In power supply systems, IEC Standard 61000-4-7 has specified conditions in the measurement of harmonics and inter-harmonics in power supply systems [116]. However, there is no standard method for measuring the sub-harmonics [115]. In practice, the period-*N* bifurcations cause the appearance of inter-harmonics and sub-harmonics. Therefore, inter-harmonics and sub-harmonics can be measured by measuring the size of the period-*N* bifurcations. Due to the high accuracy of the proposed period-*N* bifurcations identification method, we recommend the presented period-*N* bifurcations identification method as a standard method for measuring the sub-harmonics and inter-harmonics.

6.2 Prospects

(1) Fault's root cause isolation based on cumulative diagnosis approach

The cumulative diagnosis in dynamic systems may isolate a fault and find the root cause out. In the literature, researchers declare that many factors can lead to chatter, such as cutting tool wear, bearing faults, work-table faults, motor faults. In practice, it is essential to isolate the chatter's fault and find the root cause out. In our future work, we will focus on the fault's root cause isolation.

(2) Remaining useful life estimation based on cumulative diagnosis approach

The phenomenon of "self-repair" exists in dynamic systems widely. Under the condition of the "self-repair," it is difficult to identify the degree of degradations in dynamic systems reliably and estimate their remaining

useful life. The cumulative diagnosis approach can deal with this issue. In the future, the remaining useful life estimation of dynamic systems will be the other research direction.

References

[1] Peng ZK, Jackson MR, Guo LZ, Parkin RM, Meng G. Effects of bearing clearance on the chatter stability of milling process. Nonlinear Anal Real World Appl 2010;11:3577–89. https://doi.org/10.1016/j.nonrwa.2010.01.005.

[2] Haghighat M, Abdel-Mottaleb M, Alhalabi W. Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans Inf Forensics Secur 2016;11:1984–96. https://doi.org/10.1109/TIFS.2016.2569061.

[3] Williams T, Ribadeneira X, Billington S, Kurfess T. Rolling element bearing diagnostics in run-to-failure lifetime testing. Mech Syst Signal Process 2001;15:979–93. https://doi.org/10.1006/mssp.2001.1418.

[4] Qiu H, Lee J, Lin J, Yu G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. J Sound Vib 2006;289:1066–90. https://doi.org/10.1016/j.jsv.2005.03.007.

[5] Hwang W, K.S. Han. Cumulative Damage Models and Multi-Stress Fatigue Life Prediction. J Compos Mater 1986;20:125–53.

[6] Xue L. Damage accumulation and fracture initiation in uncracked ductile solids subject to triaxial loading. Int J Solids Struct 2007;44:5163–81. https://doi.org/10.1016/j.ijsolstr.2006.12.026.

[7] Adjallah KH. Cumulative diagnosis strategy for predictive maintenance decision support. 2009 Int. Conf. Comput. Ind. Eng., 2009, p. 1216–9. https://doi.org/10.1109/ICCIE.2009.5223731.

[8] Duan C, Deng C. Prognostics of Health Measures for Machines with Aging and Dynamic Cumulative Damage. IEEE/ASME Trans Mechatronics 2020;4435:1–11. https://doi.org/10.1109/tmech.2020.2995757.

[9] Zhou F, Park JH, Wen C, Hu P. Average accumulative based time variant model for early diagnosis and prognosis of slowly varying faults. Sensors 2018;18:1–27. https://doi.org/10.3390/s18061804.

[10] Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques-Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. IEEE Trans Ind Electron 2015;62:3757–67. https://doi.org/10.1109/TIE.2015.2419013.

[11] Cheng C, Wang Z, Hung W, Bukkapatnam STS, Komanduri R. Ultra-precision machining process dynamics and surface quality monitoring. Procedia Manuf 2015;1:607–18. https://doi.org/10.1016/j.promfg.2015.09.044.

[12] Honeycutt A, Schmitz TL. A study of milling surface quality during period-2 bifurcations. Procedia Manuf 2017;10:183–93. https://doi.org/10.1016/j.promfg.2017.07.046.

[13] Honeycutt A, Schmitz T. Milling bifurcations: A review of literature and experiment. J Manuf Sci Eng 2018;140:1–19. https://doi.org/10.1115/1.4041325.

[14] Cao H, Zhou K, Chen X, Zhang X. Early chatter detection in end milling based on multi-feature fusion and 3σ criterion. Int J Adv Manuf Technol 2017;92:4387–97. https://doi.org/10.1007/s00170-017-0476-x.

[15] Yang K, Wang G, Dong Y, Zhang Q, Sang L. Early chatter identification based on an optimized variational mode decomposition. Mech Syst Signal Process 2019;115:238–54. https://doi.org/10.1016/j.ymssp.2018.05.052.

[16] Kolokolov Y, Monovskaya A. Fractal approach to forming of modified bifurcation diagrams in practical applications. Int J Comput 2013;12:133–41.

[17] Zhao Y, Adjallah KH, Sava A, Wang Z. Early chatter detection using MaxEnt and SPRT. 2019 6th Int. Conf. Control. Decis. Inf. Technol., Paris, France: 2019, p. 1550–5. https://doi.org/10.1109/CoDIT.2019.8820670.

[18] Quintana G, Ciurana J. Chatter in machining processes: A review. Int J Mach Tools Manuf 2011;51:363–76. https://doi.org/10.1016/j.ijmachtools.2011.01.001.

[19] Zhu L, Liu C. Recent progress of chatter prediction, detection and suppression in milling. Mech Syst Signal Process 2020;143:1–37. https://doi.org/10.1016/j.ymssp.2020.106840.

[20] Braun S. The synchronous (time domain) average revisited. Mech Syst Signal Process 2011;25:1087–102. https://doi.org/10.1016/j.ymssp.2010.07.016.

[21] Bechhoefer E, Kingsley M. A review of time synchronous average algorithms. Annu. Conf. Progn. Heal. Manag. Soc. PHM 2009, 2009, p. 1–10.

[22] Tsai Y-H, Chang F-R, Yang W-C. GPS fault detection and exclusion using moving average filter. IEE Proc - Radar Sonar Navig 2004;151:240–7. https://doi.org/10.1049/ip-rsn.

[23] Al-Odienat AI, Al-Mbaideen AA. Optimal length determination of the moving average filter for power system applications. Int J Innov Comput Inf Control 2015;11:691–705.

[24] Hatchett RB, Brorsen BW, Anderson KB. Optimal length of moving average to forecast futures basis. J Agric Resour Econ 2010;35:18–33.

[25] Wang D, Tse PW. Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method. Mech Syst Signal Process 2015;56–57:213–29. https://doi.org/10.1016/j.ymssp.2014.10.010.

[26] Attuati S, Farina M, Boem F, Parisini T. Reducing false alarm rates in observer-based distributed fault detection schemes by analyzing moving averages. IFAC-PapersOnLine 2018;51:473–9. https://doi.org/10.1016/j.ifacol.2018.09.619.

[27] Liu YG, Hu SJ. Assembly fixture fault diagnosis using designated component analysis. J Manuf Sci Eng Trans ASME 2005;127:358–68. https://doi.org/10.1115/1.1852572.

[28] W RS. Control chart tests based on geometric moving averages. Technometrics, 1959;1:239–50.

[29] Simon DL, Bird J, Davison C, Volponi A, Iverson RE. Benchmarking gas path diagnostic methods:

A public approach. Proc ASME Turbo Expo 2008;2:325–36. https://doi.org/10.1115/GT2008-51360.

[30] Dalpiaz G, Rivola A, Rubini R. Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. Mech Syst Signal Process 2000;14:387–412. https://doi.org/10.1006/mssp.1999.1294.

[31] Aherwar A. An investigation on gearbox fault detection using vibration analysis techniques: A review. Aust J Mech Eng 2012;10:1–15. https://doi.org/10.1103/RevModPhys.82.1349.

[32] Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 2006;20:1483–510. https://doi.org/10.1016/j.ymssp.2005.09.012.

[33] Peeters C, Leclère Q, Antoni J, Lindahl P, Donnal J, Leeb S, et al. Review and comparison of tacholess instantaneous speed estimation methods on experimental vibration data. Mech Syst Signal Process 2019;129:407–36. https://doi.org/10.1016/j.ymssp.2019.02.031.

[34] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: Forecasting and Control--Fifth edition. John Wiley & Sons, Inc., Hoboken, New Jersey; 2016.

[35] Gelman A. Bayesian data analysis 3rd ed. vol. 1542. Taylor & Francis Group; 2015. https://doi.org/10.1017/CBO9781107415324.004.

[36] Särkkä S. Bayesian filtering and smoothing. Cambridge University Press; 2013. https://doi.org/10.1017/CBO9781139344203.

[37] Zhang D, Bailey AD, Djurdjanovic D. Bayesian identification of hidden Markov models and their use for condition-based monitoring. IEEE Trans Reliab 2016;65:1471–82. https://doi.org/10.1109/TR.2016.2570561.

[38] Yin Y, Huang H, Peng W, Li Y, Mi J. An E-Bayesian method for reliability analysis of exponentially distributed products with zero-failure data. Eksploat i Niezawodn – Maint Reliab 2016;18:445–9.

[39] Chen Z. Bayesian filtering: from Kalman filters to particle filters, and beyond. A J Theor Appl Stat 2003;182:1–55.

[40] Ondel O, Boutleux E, Blanco E, Clerc G. Using Kalman filter for fault diagnosis. IEEE Trans Ind Electron 2012;59:4293–300. https://doi.org/10.1109/TIE.2011.2181133.

[41] Daroogheh N, Meskin N, Khorasani K. A dual particle filter-based fault diagnosis scheme for nonlinear systems. IEEE Trans Control Syst Technol 2018;26:1317–34.

[42] Orchard ME, Vachtsevanos GJ. A particle-filtering approach for on-line fault diagnosis and failure prognosis. Trans Inst Meas Control 2009;3:221–46.

[43] Rigatos GG. Particle and Kalman filtering for state estimation and control of DC motors. ISA Trans 2009;48:62–72. https://doi.org/10.1016/j.isatra.2008.10.005.

[44] Kruschke JK, Liddell TM. The bayesian new statistics : hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. Psychon Bull Rev 2018;25:178–206.

https://doi.org/10.3758/s13423-016-1221-4.

[45] Berger JO, Brown LD. A unified conditional frequentist and bayesian test for fixed and sequential simple hypothesis testing. Ann Stat 1994;22:1787–807.

[46] Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. Ann Math Stat 1948;19:326–39.

[47] Wald A. Sequential tests of statistical hypotheses. Ann Math Stat 1945;16:117–86.

[48] Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Seq Anal, 2011;30:58–78. https://doi.org/10.1080/07474946.2011.539924.

[49] Piatyszek E, Voignier P, Graillot D. Fault detection on a sewer network by a combination of a Kalman filter and a binary sequential probability ratio test. J Hydrol 2000;230:258–68. https://doi.org/10.1016/S0022-1694(00)00213-4.

[50] Chetouani Y. A sequential probability ratio test (SPRT) to detect changes and process safety monitoring. Process Saf Environ Prot 2014;92:206–14. https://doi.org/10.1016/j.psep.2013.02.001.

[51] Zhao Y, Adjallah KH, Sava A, Wang Z. Online incipient chatter detection based on once-per-revolution sampling and dynamic threshold variant. Proc. 2019 10th IEEE Int. Conf. Intell. Data Acquis. Adv. Comput. Syst. Technol. Appl. IDAACS 2019, vol. 2, 2019. https://doi.org/10.1109/IDAACS.2019.8924402.

[52] Bin Shams MA, Budman HM, Duever TA. Fault detection, identification and diagnosis using CUSUM based PCA. Chem Eng Sci 2011;66:4488–98. https://doi.org/10.1016/j.ces.2011.05.028.

[53] Bøjer Rasmussen T, Ulrichsen SP, Nørgaard M. Use of risk-adjusted CUSUM charts to monitor 30-day mortality in danish hospitals. Clin Epidemiol 2018;10:445–56. https://doi.org/10.2147/CLEP.S157162.

[54] Page ES. Continuous inspection schemes. Biometrika, 1954;41:100–15.

[55] Castagliola P, Maravelakis PE. A CUSUM control chart for monitoring the variance when parameters are estimated. J Stat Plan Inference 2011;141:1463–78. https://doi.org/10.1016/j.jspi.2010.10.013.

[56] Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: A literature review. J Qual Technol 2006;38:349–64. https://doi.org/10.1080/00224065.2006.11918623.

[57] Li Y, Xu M, Wei Y, Huang W. Health condition monitoring and early fault diagnosis of bearings using SDF and intrinsic. IEEE Trans Instrum Meas 2016;65:1–16.

[58] Zhang Y, Zhang Y. Fault detection of non-Gaussian processes based on modified independent component analysis. Chem Eng Sci 2010;65:4630–9. https://doi.org/10.1016/j.ces.2010.05.010.

[59] Lucas JM, Crosier RB. Fast initial response for CUSUM Quality-control schemes: Give your

CUSUM A head start. Technometrics, 2000;42:102–7. https://doi.org/10.1080/00401706.2000.10485987.

[60] Westgard JO, Groth T, Aronsson T, De Verdier CH. Combined Shewhart-CUSUM control chart for improved quality control in clinical chemistry. Clin Chem 1977;23:1881–7.

[61] Abujiya MR, Riaz M, Lee MH. Improving the performance of combined Shewhart-cumulative sum control charts. Qual Reliab Eng Int 2013;29:1193–206. https://doi.org/10.1002/qre.1470.

[62] Montgomery DC. Introduction to statistical quality control (6th ed.). New York: Wiley; 2009.

[63] Lowry CA, Montgomery DC. A review of multivariate control charts. IIE Trans 1995;27:800–10. https://doi.org/10.1080/07408179508936797.

[64] Ma L, Melkote SN, Castle JB. A model-based computationally efficient method for on-line detection of chatter in milling. J Manuf Sci Eng 2013;135:031007. https://doi.org/10.1115/1.4023716.

[65] Joseph J. Pignatiello J, Runger GC. Comparisons of multivariate CUSUM charts. J Qual Technol 1990;22:173–8.

[66] Jeng J, Kang Y, Chang Y. Identification for chaos and subharmonic responses in coupled forced Duffing's oscillators. J Mech 2011;27:139–47. https://doi.org/10.1017/jmech.2011.15.

[67] Mcfadden PD. A revised model for the extraction of periodic waveforms by time domain averaging. Mech Syst Signal Process 1987;1:83–95.

[68] Huang F, Sava A, Adjallah KH, Wang Z. Bearings degradation monitoring indicator based on segmented hotelling t square and piecewise linear representation. Proc 2018 IEEE Int Conf Mechatronics Autom ICMA 2018 2018:1389–94. https://doi.org/10.1109/ICMA.2018.8484434.

[69] Honeycutt A, Schmitz T. A numerical and experimental investigation of period-n bifurcations in milling. J Manuf Sci Eng 2017;139:1–11. https://doi.org/10.1115/1.4034138.

[70] Altintas Y. In-process detection of tool breakages using time series monitoring of cutting forces. Int J Mach Tools Manuf 1988;28:157–72. https://doi.org/10.1016/0890-6955(88)90027-2.

[71] Zhao Y, Adjallah KH, Sava A, Wang Z. MaxEnt feature-based reliability model method for real-time detection of early chatter in high-speed milling. ISA Trans 2020:109231. https://doi.org/10.1016/j.econlet.2020.109231.

[72] Kolokolov Y, Monovskaya A. From modifications of experimental bifurcation diagrams to operating process stability margin. Int J Bifurc Chaos 2013;23:1–20. https://doi.org/10.1142/S0218127413300243.

[73] Cai B, Huang L, Xie M. Bayesian networks in fault diagnosis. IEEE Trans Ind Informatics 2017;13:2227–40. https://doi.org/10.1109/TII.2017.2695583.

[74] Mechraoui A, Medjaher K, Zerhouni N. Bayesian based fault diagnosis: Application to an electrical motor. IFAC Proc. Vol., vol. 41, 2008, p. 7381–6. https://doi.org/10.3182/20080706-5-KR-1001.2841.

[75] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mech Syst Signal Process 2020;138:106587. https://doi.org/10.1016/j.ymssp.2019.106587.

[76] Gangsar P, Tiwari R. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. Mech Syst Signal Process 2020;144:106908. https://doi.org/10.1016/j.ymssp.2020.106908.

[77] Nguyen NT, Lee HH, Kwon JM. Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor. J Mech Sci Technol 2008;22:490–6. https://doi.org/10.1007/s12206-007-1036-3.

[78] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mech Syst Signal Process 2020;138:106587. https://doi.org/10.1016/j.ymssp.2019.106587.

[79] Gangsar P, Tiwari R. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. Mech Syst Signal Process 2020;144:106908. https://doi.org/10.1016/j.ymssp.2020.106908.

[80] Khawaja T, Vachtsevanos G, Wu B. Reasoning about Uncertainty in Prognosis: A Confidence Prediction Neural Network. NAFIPS 2005 - 2005 Annu. Meet. North Am. Fuzzy Inf. Process. Soc. Reason., 2005, p. 7–12.

[81] Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Ann - Manuf Technol 2010;59:717–39. https://doi.org/10.1016/j.cirp.2010.05.010.

[82] Jemielniak K. Commercial tool condition monitoring systems. Int J Adv Manuf Technol 1999:711–21. https://doi.org/10.1007/s001700050123.

[83] Yue C, Gao H, Liu X, Liang S, Wang L. A review of chatter vibration research in milling. Chinese J Aeronaut 2019;32:215–42. https://doi.org/10.1016/j.cja.2018.11.007.

[84] Ferrero A, Ottoboni R. High-accuracy Fourier analysis based on synchronous sampling techniques. IEEE Trans Instrum Meas 1992;41:780–5.

[85] Kakinuma Y, Sudo Y, Aoyama T. Detection of chatter vibration in end milling applying disturbance observer. CIRP Ann - Manuf Technol 2011;60:109–12. https://doi.org/10.1016/j.cirp.2011.03.080.

[86] Kolokolov Y V., Monovskaya A V. Preventive diagnosis of pulse power converter dynamics. Autom Remote Control 2009;70:1228–42. https://doi.org/10.1134/S0005117909070133.

[87] Honeycutt A, Schmitz TL. Milling stability interrogation by subharmonic sampling. J Manuf Sci Eng 2016;139:041009. https://doi.org/10.1115/1.4034894.

[88] Schmitz TL, Medicus K, Dutterer B. Exploring once-per-revolution audio signal variance as a chatter indicator. Mach Sci Technol 2002;6:215–33. https://doi.org/10.1081/MST-120005957.

[89] Davies MA, Pratt JR, Dutterer BS, Burns TJ. The stability of low radial immersion milling. CIRP

Ann 2000;49:37–40.

[90] Honeycutt A, Schmitz TL. Experimental validation of period-n bifurcations in milling. Procedia Manuf 2016;5:362–74. https://doi.org/10.1016/j.promfg.2016.08.031.

[91] Schmitz TL. Chatter recognition by a statistical evaluation of the synchronously sampled audio signal. J Sound Vib 2003;262:721–30. https://doi.org/10.1016/S0022-460X(03)00119-6.

[92] Honeycutt A, Schmitz TL. A new metric for automated stability identification in time domain milling simulation. J Manuf Sci Eng 2016;138:1–7. https://doi.org/10.1115/1.4032586.

[93] Cao H, Zhou K, Chen X, Zhang X. Early chatter detection in end milling based on multi-feature fusion and 3σ criterion. Int J Adv Manuf Technol 2017;92:4387–97. https://doi.org/10.1007/s00170-017-0476-x.

[94] Jia G, Wu B, Hu Y, Xie F, Liu A. A synthetic criterion for early recognition of cutting chatter. Sci China Technol Sci 2013;56:2870–6. https://doi.org/10.1007/s11431-013-5360-9.

[95] Hynynen KM, Ratava J, Lindh T, Rikkonen M, Ryynänen V, Lohtander M, et al. Chatter detection in turning processes using coherence of acceleration and audio signals. J Manuf Sci Eng 2014;136:044503. https://doi.org/10.1115/1.4026948.

[96] Tansel IN, Li M, Demetgul M, Bickraj K, Kaya B, Ozcelik B. Detecting chatter and estimating wear from the torque of end milling signals by using Index Based Reasoner (IBR). Int J Adv Manuf Technol 2012;58:109–18. https://doi.org/10.1007/s00170-010-2838-5.

[97] Yao Z, Mei D, Chen Z. On-line chatter detection and identification based on wavelet and support vector machine. J Mater Process Technol 2010;210:713–9. https://doi.org/10.1016/j.jmatprotec.2009.11.007.

[98] Rusinek R, Pawel L, Krzysztof K, Kruszynski B, Warminski J. Chatter identification methods on the basis of time series measured during titanium superalloy milling. Int J Mech Sci 2015;99:196–207. https://doi.org/10.1016/j.ijmecsci.2015.05.013.

[99] Aslan D, Altintas Y. On-line chatter detection in milling using drive motor current commands extracted from CNC. Int J Mach Tools Manuf 2018;132:64–80. https://doi.org/10.1016/j.ijmachtools.2018.04.007.

[100] Liu C, Zhu L, Ni C. Chatter detection in milling process based on VMD and energy entropy. Mech Syst Signal Process 2018;105:169–82. https://doi.org/10.1016/j.ymssp.2017.11.046.

[101] Van Dijk NJM, Doppenberg EJJ, Faassen RPH, Van de Wouw N, Oosterling JAJ, Nijmeijer H. Automatic in-process chatter avoidance in the high-speed milling process. J Dyn Syst Meas Control 2010;132:031006. https://doi.org/10.1115/1.4000821.

[102] Messaoud A, Weihs C. Monitoring a deep hole drilling process by nonlinear time series modeling. J Sound Vib 2009;321:620–30. https://doi.org/10.1016/j.jsv.2008.10.028.

[103] Caliskan H, Kilic ZM, Altintas Y. On-line energy-based milling chatter detection. J Manuf Sci Eng

Trans ASME 2018;140:1–12. https://doi.org/10.1115/1.4040617.

[104] Du R, Elbestawi MA, Ullagaddi BC. Chatter detection in milling based on the probability distribution of cutting force signal. Mech Syst Signal Process 1992;6:345–62. https://doi.org/10.1016/0888-3270(92)90036-I.

[105] Wang L, Liang M. Chatter detection based on probability distribution of wavelet modulus maxima. Robot Comput Integr Manuf 2009;25:989–98. https://doi.org/10.1016/j.rcim.2009.04.011.

[106] Govekarl E, GradiSek J, Kalveram M, Insperger T, Weinert K, Stepan G, et al. On stability and dynamics of milling at small radial immersion. CIRP Ann, 2005;54:357–62.

[107] Insperger T, Mann BP, Surmann T, Stépán G. On the chatter frequencies of milling processes with runout. Int J Mach Tools Manuf 2008;48:1081–9. https://doi.org/10.1016/j.ijmachtools.2008.02.002.

[108] Cao H, Lei Y, He Z. Chatter identification in end milling process using wavelet packets and Hilbert-Huang transform. Int J Mach Tools Manuf 2013;69:11–9. https://doi.org/10.1016/j.ijmachtools.2013.02.007.

[109] Deng C, Miao J, Ma Y, Wei B, Feng Y. Reliability analysis of chatter stability for milling process system with uncertainties based on neural network and fourth moment method. Int J Prod Res 2019;58:2732–50. https://doi.org/10.1080/00207543.2019.1636327.

[110] Mallat S, Hwang WL. Singularity detection and processing with wavelets. IEEE Trans Inf Theory 1992;38:617–43.

[111] Albertelli P, Braghieri L, Torta M, Monno M. Development of a generalized chatter detection methodology for variable speed machining. Mech Syst Signal Process 2019;123:26–42. https://doi.org/10.1016/j.ymssp.2019.01.002.

[112] Cao H, Zhang X, Chen X. The concept and progress of intelligent spindles: A review. Int J Mach Tools Manuf 2017;112:21–52. https://doi.org/10.1016/j.ijmachtools.2016.10.005.

[113] Kolokolov Y V., Monovskaya A V., Adjallah KH. PWM energy converters: Fractal method of dynamics forecasting in practical application. IEEE Trans Energy Convers 2009;24:483–92. https://doi.org/10.1109/TEC.2009.2016043.

[114] Leonowicz Z. Analysis of sub-harmonics in power systems. 2010 9th Conf Environ Electr Eng EEEIC 2010 2010:125–7. https://doi.org/10.1109/EEEIC.2010.5489951.

[115] Barros J, De Apraiz M, Diego RI. Measurement of subharmonics in power voltages. 2007 IEEE Lausanne Power Tech 2007:1736–40. https://doi.org/10.1109/PCT.2007.4538578.

[116] Diego RI, Barros J. Subharmonic measurement using DFT and Wavelet-Packet Transform in an IEC extended framework. Measurement 2010;43:1603–8. https://doi.org/10.1016/j.measurement.2010.09.012.

[117] Kolokolov Y, Monovskaya A. Fractal regularities of sub-harmonic motions perspective for pulse dynamics monitoring. Chaos, Solitons and Fractals 2005;23:231–41.

https://doi.org/10.1016/j.chaos.2004.04.012.

[118] Davies MA, Dutterer B, Pratt JR, Schaut AJ, Bryan JB. On the dynamics of high-speed milling with long, slender endmills. CIRP Ann - Manuf Technol 1998;47:55–60. https://doi.org/10.1016/S0007-8506(07)62784-X.

[119] Faulkner ND, Mestre EVI. Subharmonic Sampling for the Measurement of Short-Term Stability of Microwave Oscillators. IEEE Trans Instrum Meas 1983;32:208–13. https://doi.org/10.1109/TIM.1983.4315043.

[120] Wang F, Yi H, Zhuo F, Xiong L, Liu X, Zhu M. A quantitative evaluation and comparison of harmonic elimination algorithms based on moving average filter and delayed signal cancellation in phase synchronization applications. J Power Electron 2016;16:717–30. https://doi.org/10.6113/jpe.2016.16.2.717.

[121] Freijedo FD, Doval-Gandoy J, Lopez O, Cabaleiro J. Harmonic identification methods based on moving average filters for active power filters. Conf Rec - IAS Annu Meet (IEEE Ind Appl Soc 2008:1–6. https://doi.org/10.1109/08IAS.2008.272.

[122] Weinberg A, Liu B. Discrete time analyses of nonuniform sampling first- and second-order digital phase lock loops. IEEE Trans Commun 1974;22:123–37. https://doi.org/10.1109/TCOM.1974.1092168.

[123] Ferrero A, Ottoboni R. Synchronous sampling techniques. Instrumentation 1992;41:780–5.

[124] Borghesani P, Pennacchi P, Chatterton S, Ricci R. The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery. Mech Syst Signal Process 2014;44:118–33. https://doi.org/10.1016/j.ymssp.2013.03.026.

[125] Braun S. The synchronous (time domain) average revisited. Mech Syst Signal Process 2011;25:1087–102. https://doi.org/10.1016/j.ymssp.2010.07.016.

[126] Li H, Zhang Y, Zheng H. Bearing fault detection and diagnosis based on order tracking and Teager-Huang transform. J Mech Sci Technol 2010;24:811–22. https://doi.org/10.1007/s12206-009-1211-9.

[127] Luo H, Qiu H, Ghanime G, Hirz M, van der Merwe G. Synthesized Synchronous Sampling Technique for Differential Bearing Damage Detection. J Eng Gas Turbines Power 2010;132. https://doi.org/10.1115/1.4000092.

[128] Lu S, Yan R, Liu Y, Wang Q. Tacholess speed estimation in order tracking: A review with application to rotating machine fault diagnosis. IEEE Trans Instrum Meas 2019;68:1–18. https://doi.org/10.1109/TIM.2019.2902806.

[129] Jaynes ET. Information theory and statistical mechanics. Inf Theory Stat Phys 1957;106:620–30. https://doi.org/10.1103/PhysRev.106.620.

[130] Jaynes ET. Information entropy and statistical mechanics II. Phys Rev 1957;108:171–90. https://doi.org/10.1103/PhysRev.106.620.

[131] Kang HY, Kwak BM. Application of maximum entropy principle for reliability-based design optimization. Struct Multidiscip Optim 2009;38:331–46. https://doi.org/10.1007/s00158-008-0299-3.

[132] Li H, Wen D, Lu Z, Wang Y, Deng F. Identifying the probability distribution of fatigue life using the maximum entropy principle. Entropy 2016;18. https://doi.org/10.3390/e18040111.

[133] Li C, Wang S, Wang W. Reliability analysis of free jet scour below dams. Entropy 2012;14:2578–88. https://doi.org/10.3390/e14122578.

[134] Li C, Wang W, Wang S. Maximum-entropy method for evaluating the slope stability of earth dams. Entropy, 2012;14:1864–76. https://doi.org/10.3390/e14101864.

[135] Rajan A, Kuang YC, Ooi MPL, Demidenko S, Carstens H. Moment-constrained maximum entropy method for expanded uncertainty evaluation. IEEE Access 2018;6:4072–82. https://doi.org/10.1109/ACCESS.2017.2787736.

[136] Luigi P, Inverardi N. Maximum Entropy Density Estimation from Fractional Moments. Commun Stat Theory Methods 2003;32:327–45. https://doi.org/10.1081/STA-120018189.

[137] Zhang X, Pandey MD. Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method. Struct Saf 2013;43:28–40. https://doi.org/10.1016/j.strusafe.2013.03.001.

[138] Zhang X, Min Y, Ghee C. Maximum entropy distribution with fractional moments for reliability analysis. Struct Saf 2020;83:101904. https://doi.org/10.1016/j.strusafe.2019.101904.

[139] Liu Y, Meng L lin, Liu K, Zhang Y min. Chatter reliability of milling system based on first-order second-moment method. Int J Adv Manuf Technol 2016;87:801–9. https://doi.org/10.1007/s00170-016-8523-6.

[140] Rausand M, Hoyland A. System reliability theory: models, statistical methods, and applications. 2nd Ed. John Wiley & Sons, Inc.; 2004.

[141] Harbitz A. An efficient sampling method for probability of failure calculation. Struct Saf 1986;3:109–15. https://doi.org/10.1016/0167-4730(86)90012-3.

[142] Huang X, Zhang Y, Lv C. Probabilistic analysis of dynamic stability for milling process. Nonlinear Dyn 2016;86:2105–14. https://doi.org/10.1007/s11071-016-3019-3.

[143] Favretti M. Remarks on the maximum entropy principle with application to the Maximum Entropy Theory of Ecology. Entropy 2018;20. https://doi.org/10.3390/e20010011.

[144] Li Y, Zhao G. Reliability analysis of structures based on maximum entropy theory. J Dalian Univ Technol 1992;32:455–9.

[145] Chien TT. An adaptiave technique for a redundant-sensor navigation system. Massachusetts Institute of Technology, 1972.

[146] Cao H, Zhang X, Chen X. The concept and progress of intelligent spindles: A review. Int J Mach

Tools Manuf 2017;112:21–52. https://doi.org/10.1016/j.ijmachtools.2016.10.005.

[147] Wu L, Yao B, Peng Z, Guan Y. An adaptive threshold algorithm for sensor fault based on the grey theory. Adv Mech Eng 2017;9:1–7.

[148] Wiercigroch M. Modelling of dynamical systems with motion dependent discontinuities. Chaos, Solitons and Fractals 2000;11:2429–42. https://doi.org/10.1016/S0960-0779(00)00032-1.

[149] Smith S, Tlusty J. An overview of modeling and simulation of the milling process. J Eng Ind 1991;113:169. https://doi.org/10.1115/1.2899674.

[150] Schmitz TL, Smith KS. Machining dynamics: Frequency response to improved productivity. New York: Springer; 2009.

[151] Zellner A, A.Highfield R. Calculation of maximum entropy distributions and approximation of marginalposterior distributions. J Econom 1988;37:195–209.

[152] Khudanpur S, Wu J. Maximum entropy techniques for exploiting syntactic, semantic and collocational dependencies in language modeling. Comput Speech Lang 2000;14:355–72. https://doi.org/10.1006/csla.2000.0149.

[153] Pincus SM, Gladstone IM, Ehrenkranz RA. A regularity statistic for medical data analysis. J Clin Monit 1991;7:335–45.

[154] Richman JS, Lake DE, Moorman JR. Sample entropy. Methods Enzymol 2004;384:172–84. https://doi.org/10.1016/S0076-6879(04)84011-4.

[155] https://ww2.mathworks.cn/matlabcentral/fileexchange/32427-fast-approximate-entropy?s_tid=srchtitle n.d.

[156] Al-Sharhan S, Karray F, Gueaieb W, Basir O. Fuzzy entropy: A brief survey. IEEE Int Conf Fuzzy Syst 2001;3:1135–9. https://doi.org/10.1109/fuzz.2001.1008855.

[157] https://datashare.is.ed.ac.uk/handle/10283/2099 n.d.

[158] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomed Signal Process Control 2014;14:19–29. https://doi.org/10.1016/j.bspc.2014.06.009.

[159] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 2009;01:1–41. https://doi.org/10.1142/S1793536909000047.

[160] Xie L, Zeng J, Kruger U, Wang X, Geluk J. Fault detection in dynamic systems using the Kullback–Leibler divergence. Control Eng Pract 2015;43:39–48. https://doi.org/10.1016/j.conengprac.2015.05.010.

[161] Wang CC. Bifurcation analysis of an aerodynamic journal bearing system considering the effect of stationary herringbone grooves. Chaos, Solitons and Fractals 2007;33:1532–45. https://doi.org/10.1016/j.chaos.2006.03.011.

[162] Kolokolov Y, Monovskaya A. Fractal approach, bifurcation poker and SUC--logic for nonlinear

dynamics forecasting. Int J Bifurc Chaos 2013;23:1–18. https://doi.org/10.1142/S0218127413502015.

[163] Kolokolov Y, Monovskaya A. Estimating the uncertainty of the behavior of a PWM power converter by analyzing a set of experimental bifurcation diagrams. Int J Bifurc Chaos 2013;23:1–14. https://doi.org/10.1142/S0218127413500636.

[164] Fellouris G, Tartakovsky AG. Nearly minimax one-sided mixture-based sequential tests. Seq Anal Des Methods Appl 2012;31:297–325. https://doi.org/10.1080/07474946.2012.694346.

[165] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng QN, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc a-Mathematical Phys Eng Sci 1998;454:903–95.ss https://doi.org/10.1098/rspa.1998.0193.

Appendix A

A-1 Sequential probability ratio test

SPRT is a powerful method for hypotheses tests (step by step) [47][46]. This method was used, for instance, for detecting the change point or status in a wide range of fields, such as medicine [48], sewer network [49], and chemical industry [50]. As classical hypothesis testing, SPRT also needs two hypotheses at least: the null hypothesis, denoted H_0, corresponding to the normal mode, and the alternative hypothesis, denoted H_1, corresponding to the abnormal mode.

$$H_0 : p = p_0 , \tag{A-1}$$

$$H_1 : p = p_1 . \tag{A-2}$$

The log-likelihood ratio (LLR), defined as the logarithm of the likelihood ratio of the above two hypotheses, is given by the formula:

$$\log \Lambda(x_i) = \log\left(\frac{p(x_i / H_1)}{p(x_i / H_0)} \right). \tag{A-3}$$

When a new sample x_n is collected, the cumulative log-likelihood ratio (CLLR) is calculated as

$$S_n = S_{n-1} + \log \Lambda(x_n) = \log\left(\frac{p(x_1, x_2, \ldots, x_n / H_1)}{p(x_1, x_2, \ldots, x_n / H_0)} \right), \tag{A-4}$$

where, n=1, 2, 3, …, and S_0=0.

The SPRT considers two decision risks, i.e., α and β. These two decision risks can be preset according to the reliability requirement or risks acceptance, which allow setting the decision threshold values as following:

- If the $S_n \leq a$, accept H_0 and stop the detection
- If the $S_n \geq b$, accept H_1 and stop the detection
- If the $a < S_n < b$, continue the detection until either a or b is reached,

where two threshold values a and b depend on the two decision risks,

$$a = \log \frac{\beta}{1-\alpha}, \tag{A-5}$$

$$b = \log \frac{1-\beta}{\alpha}. \tag{A-6}$$

For decision, the CLLR S_n is compared to both thresholds values a and b. The process can be supervised in a recursive and real-time manner.

A-2 Modified sequential probability ratio test

The original SPRT detects the normal process and abnormal process reliably [46]. However, in some situations, the detection of only one process (normal or abnormal) is interesting. For instance, in the detection of the chatter (abnormal process), only the detection of the chatter is interesting. In this situation, the detection of the chatter-free state (normal process) may increase detection delay in the chatter detection [145]. To detect early-chatter as soon as possible, we propose to use a modified SPRT (M-SPRT), which does not need to detect the unconcerned process and reduces the detection delay.

Since M-SPRT only detects either normal process or abnormal process, it is also called as one-sided SPRT [164]. As the original SPRT, the M-SPRT also assumes two hypotheses expressed in equations (A-1) and (A-2).

For receiving data x_i, its probability from the normal process will be $p(x_i / H_0)$. Otherwise, its probability will be $p(x_i / H_1)$. The CLLR is calculated according to equation (A-4).

To avoid the detection of the chatter-free process, we modify the CLLR as follows based on the distribution of the feature (or indicator):

$$S_n = Max\{0, S_n\}, \tag{A-7}$$

$$\text{or} \quad S_n = Min\{0, S_n\}. \tag{A-8}$$

The CLLR S_n is reset to $Max\{0, S_n\}$, as shown in (A-7), if the mean of the feature in an abnormal process is higher than that in the normal process. Only the upper threshold corresponding to the abnormal process may be reached in the detection process. The corresponding stopping rule is as follows:

- Accept H_1 if $S_n \geq b$,
- Continue monitoring if $S_n < b$.

Similarly, the CLLR S_n is reset to $Min\{0, S_n\}$, as shown in (A-8), if the mean of the feature in an abnormal process is less than that in the normal process. Only a lower threshold corresponding to the abnormal process may be crossed. The corresponding stopping rule is expressed as follows:

- Accept H_1 if $S_n \leq a$,
- Continue monitoring if $S_n > a$,

where a and b represent the lower threshold and upper threshold, respectively. Two threshold values are calculated according to equations (A-5) and (A-6).

From the procedure of M-SPRT, it can be seen that M-SPRT is very simple to implement for recognizing the abnormal process.

Appendix B

B-1 ICEEMDAN

The improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) method was introduced by Colominas et al. [158] to eliminate the residual noises in the intrinsic mode functions (IMFs) and to address the issue of "spurious" modes in the early decomposition stages of the original complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), where the white noise is added in the first decomposition stage. In ICEEMDAN, peculiar noises are added at all decomposition stages, and each IMF is defined as a difference between the last residual $r_{i-1}(k)$ and the current residual $r_i(k)$:

$$c_i(k) = r_{i-1}(k) - r_i(k),\ i = 1,2,3,\cdots n, \text{ with } r_0 = x(k). \tag{B-1}$$

Let $E_i(\omega(k))$ denote the i^{th} IMF of white noises decomposed using the empirical mode decomposition (EMD) method, where $E_i(\cdot)$ is the operator used to produce the i^{th} mode decomposed by EMD and $\omega(k)$ is a white noise signal with zero mean and unit variance. In order to conveniently describe the ICEEMDAN algorithm, one considers the local mean operator $\bar{M}(\cdot)$ applied to the signal to calculate the next residue at each stage. The ICEEMDAN algorithm is described as it follows:

(1) Construct the mixed-signal $x_m(k)$ by adding $E_1(\omega_m(k))$ to the original signal $x(k)$:

$$x_m(k) = x(k) + \beta_1 std(x(k)) / std(E_1(\omega_m(k)))E_1(\omega_m(k)), \tag{B-2}$$

where $\omega_m(k)$ represents the m^{th} added white noise, β_1 represents the noise magnitude parameter and $std(\bullet)$ stands for an estimating operator of the standard deviation of a signal.

(2) Calculate the local mean $\bar{M}(x_m(k))$ using EMD and obtain the first residual by an average of M realizations:

$$r_1(k) = \frac{1}{M}\sum_{m=1}^{M}\bar{M}(x_m(k)). \tag{B-3}$$

Then, calculate the first IMF $c_1(k) = x(k) - r_1(k)$;

(3) Compute the second IMF $c_2(k) = r_1(k) - r_2(k)$, with $r_2(k) = \frac{1}{M}\sum_{m=1}^{M}\bar{M}(r_1(k) + \beta_2 std(r_1(k))E_2(\omega_m(k)))$;

(4) Similarly, the i^{th} IMF: $c_i(k) = r_{i-1}(k) - r_i(k)$, with $r_i(k) = \frac{1}{M}\sum_{m=1}^{M}\bar{M}(r_{i-1}(k) + \beta_i std(r_{i-1}(k))E_i(\omega_m(k)))$, $i = 3,4,\cdots,n$.

(5) Repeat step (4) for i+1.

The decomposition process stops when the stopping criterion is satisfied according to the EMD method [165]. Through the decomposition process, the original signal is decomposed into *n* IMFs and a residue. Consequently, the original signal can be expressed as

$$x(k) = \sum_{i=1}^{n} c_i(k) + r_n . \tag{B-4}$$

The noise parameter β_i can be selected according to the required signal to noise ratio (SNR) at each stage. In order to highlight the influence of both intermittent wave amplitude and sampling frequency ratio on the ICEEMDAN mode mixing alleviation performance, we consider a signal (B-7) made of two components: the first component is an intermittent high-frequency signal $x_1(t)$ (B-5), and the second component is a low-frequency sinusoidal signal $x_2(t)$ (B-6).

$$x_1(t) = \begin{cases} a_0 \sin(2\pi \cdot 85(t-1.88)) & 1.88 \le t \le 2.12 \\ a_0 \sin(2\pi \cdot 85(t-8.38)) & 8.38 \le t \le 8.62 \\ a_0 \sin(2\pi \cdot 85(t-11.88)) & 11.88 \le t \le 12.12 \\ a_0 \sin(2\pi \cdot 85(t-17.88)) & 17.88 \le t \le 18.12 \\ 0 & others \end{cases}, \tag{B-5}$$

$$x_2(t) = \sin(2\pi \cdot 10t + \frac{\pi}{2}), \tag{B-6}$$

$$x(t) = x_1(t) + x_2(t) . \tag{B-7}$$

This signal is decomposed using ICEEMDAN. We set the noise magnitude parameter to 0.2 and the ensemble number to 100 as used in [158]. In order to show the influence of the sampling frequency ratio between the sampling frequency and maximum frequency in the signal, the abscissa is expressed using a sampling point. The digital signal and two components are denoted as $x(k)$, $x_1(k)$ and $x_2(k)$, where k represents the sampling point. The decomposition results of the signal with the intermittent component amplitude $a_0 = 0.3$ and $a_0 = 0.7$ are illustrated in Figure B-1(a) and (b), respectively.

The desired decomposition performance is obtained when the amplitude of the intermittent wave is $a_0 = 0.3$ as shown in Figure B-1(a). However, a significant mode mixing occurs when the amplitude of the intermittent wave is $a_0 = 0.7$ as shown in Figure B-1(b). Using the same setting for the noise magnitude parameter and the sampling frequency ratio, one can see that the ICEEMDAN capability of alleviating mode mixing degrades when the intermittent wave amplitude is relatively large. Besides, we increased the sampling frequency ratio by 25% (from 200/85 to 250/85) and decomposed it again. The decomposition result is displayed in Figure B-2. The mode mixing is alleviated again. This observation enabled us to think that the ICEEMDAN method might be able to alleviate the mode mixing caused by the relatively large amplitude of the intermittent wave, through a suitably selected sampling frequency ratio. Therefore, in the

next section, we study the influence of both the intermittent wave amplitude and the sampling frequency ratio on mode mixing alleviation performance of the ICEEMDAN method.

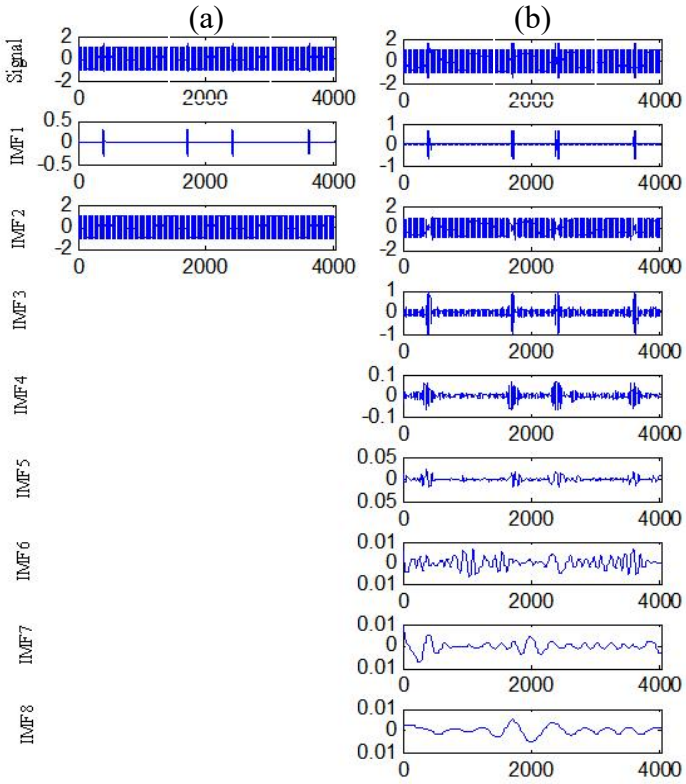


Figure B-1 First eight IMFs decomposed by ICEEMDAN, (a): $a_0 = 0.3$ and sampling frequency ratio $R_s = 200/85$; (b): $a_0 = 0.7$ and sampling frequency ratio $R_s = 200/85$.

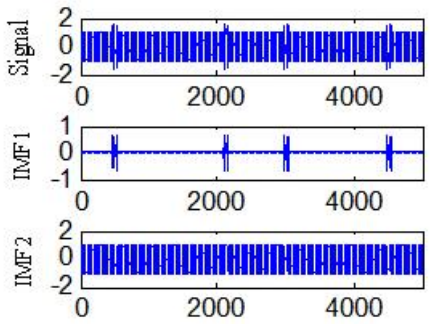


Figure B-2 Decomposition results of ICEEMDAN ($a_0 = 0.7$ and sampling frequency ratio $R_s = 250/85$).

B-2 Alleviation performance analysis of mode mixing

B-2.1 Evaluation parameters

For the performance assessment, we used in this paper the root relative squared error (RRSE) and the mean absolute error (MAE). The IMF obtained by decomposition yields the recovered signal $b(k)$, while the signal $a(k)$ stands for the reference. On the one hand, $b(k)$ is the closest IMF to the reference $a(k)$ in the RRSE performance indicator, which is defined equation (9):

$$RRSE = \sqrt{\frac{\sum_{k=1}^{S}(a(k)-b(k))^2}{\sum_{k=1}^{S}a^2(k)}}, \tag{B-9}$$

where S represents the total number of the indicators' data points.

In the MAE performance indicator assessment equation (B-10), $b(k)$ is the closest IMF to the reference $a(k)$.

$$MAE = \frac{1}{S}\sum_{k=1}^{S}|a(k)-b(k)|. \tag{B-10}$$

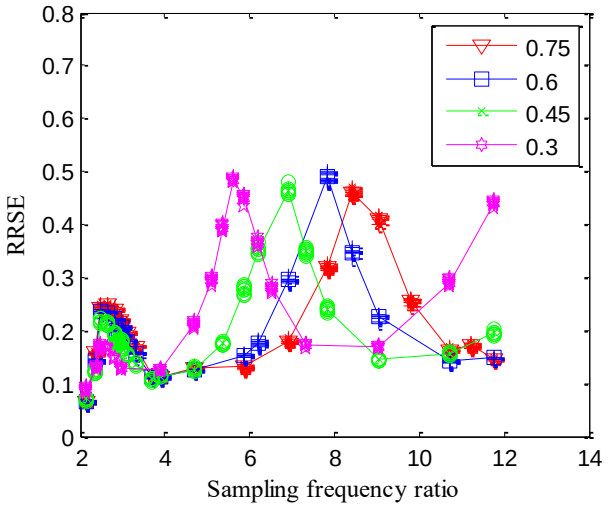
The RRSE and MAE evaluate the decomposition accuracy of the signal. If the RRSE and MAE are small or close to zero, it means that the corresponding IMF is close to the original component, and the performance of the mode mixing alleviation is high. Conversely, the mode mixing alleviation performance is imperfect.

B-2.2 Influence of both intermittent wave amplitude and sampling frequency ratio

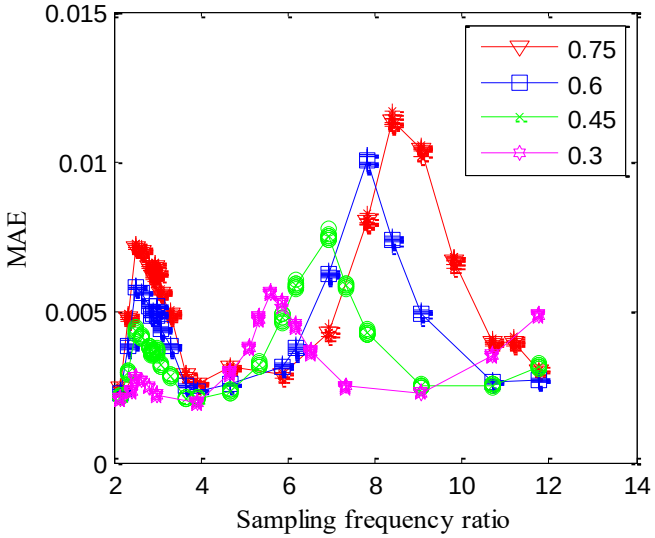
From the decomposition results of the signal given in the last section, it can be seen that the intermittent wave amplitude and sampling frequency ratio influence the ICEEMDAN mode mixing alleviation performance. In order to alleviate the mode mixing and obtain the optimal extraction results, an appropriate sampling frequency ratio needs to be selected for signals with intermittent wave components. Therefore, in this section, we analyze the influence of both intermittent wave amplitude and sampling frequency ratio on mode mixing alleviation performance of the ICEEMDAN.

We consider the signal presented in equation (B-7) to analyze the mode mixing alleviation performance of the ICEEMDAN. Four amplitudes, 0.3, 0.45, 0.6, and 0.75, respectively, of the intermittent wave signal with high-frequency and different settings for the sampling frequencies, were selected. According to the Nyquist-Shannon sampling theorem, we can reconstruct the signal when the sampling frequency is more than twice the maximum frequency of the original signal. Thus, we set the minimum sampling frequency greater than two times the maximum frequency of the original signal, i.e., the sampling frequency ratio is greater than 2. A high sampling frequency will lead to higher computation cost. Consequently, we set the sampling frequency ratio from 2.14 to 11.76. Every experiment was run ten times decomposition by ICEEMDAN while considering the random noises. The average of the ten results is used to evaluate the alleviation performance for mode mixing. The decomposition results by the ICEEMDAN are displayed in Figure B-3 and Figure B-4. In these two figures, the abscissae correspond to the frequency ratio, while the RRSE and MAE are displayed on the ordinates. Each curve relates to a given amplitude of the intermittent sinusoidal wave component. Figure B-3(a) and (b) show the RRSE and the MAE of the intermittent sinusoidal wave components extracted by the ICEEMDAN. The intermittent wave amplitude and the sampling frequency

ratio dramatically influence the RRSE and MAE of the intermittent sinusoidal wave component. For each amplitude setting, one can observe two peaks in the whole range of the sampling frequency ratio.



(a) RRSE of the intermittent wave component.



(b) MAE of intermittent wave component.

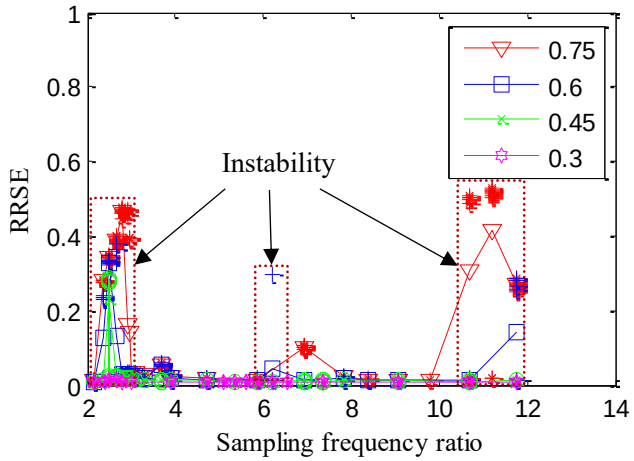
Figure B-3 Variation of the RRSE and MAE as a function of sampling frequency ratio for the intermittent wave component: a) RRSE, b) MAE.

The first peak is caused by the mode mixing between the two components contained in the signal (B-7). It increases with the increase of the amplitude of the intermittent component for both RRSE and MAE, and it is located at the same sampling frequency ratio. This phenomenon means that significant mode mixing can occur with a high amplitude of the intermittent component.

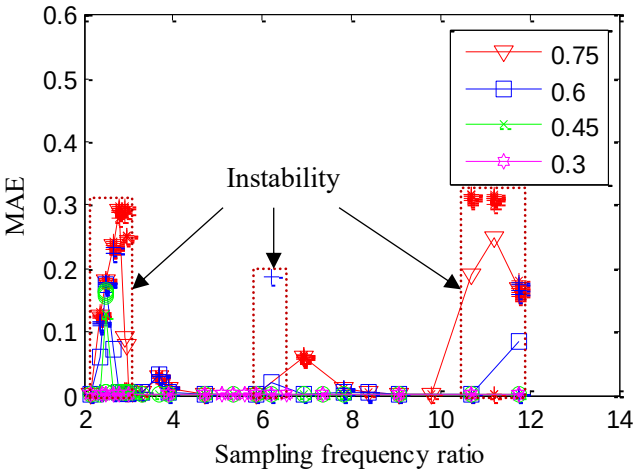
The second peak is caused by the mode mixing between the intermittent wave component and the "spurious" mode(s). The sampling frequency, associated with this peak, increases with the amplitude of the intermittent component. For the RRSE of the intermittent wave component, the second peak value has few changes with the increase of the amplitude of the first component. The maximum RRSE of the first component is not

decreased by adjusting the sampling frequency ratio, but the alleviation performance of mode mixing can be improved by selecting a suitable sampling frequency ratio. For MAE, the second peak value increases approximate linearly with the amplitude of the intermittent wave component.

The RRSE and MAE of the second component are displayed in Figure B-4(a) and (b). The decomposition performance is improved when the intermittent wave signal amplitude decreases. For the intermittent wave amplitude of 0.3, the extraction results are ideal in the whole range of sampling frequency. However, for relative high amplitudes of the intermittent wave components (amplitude of 0.45, 0.6, or 0.75 for instance), mode mixing appears at a sampling frequency ratio. It must be pointed out that the decomposition instability occurs at relatively high amplitude and peculiar sampling frequency ratio, as shown in brown dotted boxes of Figure B-4(a) and (b). This decomposition instability can be explained due to the relatively small amplitude noise. This noise cannot change absolutely extrema distribution of the signal under the conditions of the relatively large amplitude of the intermittent wave and the peculiar sampling frequency ratio.



(a) RRSE of the sinusoidal signal.



(b) MAE of the sinusoidal signal.

Figure B-4 Variation of the RRSE and MAE as a function of sampling frequency ratio: (a) RRSE, (b) MAE.

According to the above analysis, the intermittent wave amplitude and the sampling frequency ratio dramatically affect the alleviation performance of mode mixing. To alleviate mode mixing and obtain optimal extraction results with the ICEEMDAN, we should select an appropriate sampling frequency considering the amplitude of the intermittent signal.