5. L'architecture décentralisée n'a pas entièrement répondu aux attentes et conduit à s'interroger sur la pertinence du signal de prix

Le mécanisme de capacité français, lancé pour première année de livraison au 1er janvier 2017, repose sur plusieurs principes fondamentaux comme l'engagement de l'ensemble de la capacité disponible (aspect *capacity-wide*) par opposition à une réserve stratégique, son fonctionnement via un marché (aspect *market-based*) par opposition à une prime administrée, ou encore son caractère décentralisé. L'architecture du mécanisme de capacité est dite décentralisée car elle repose sur une obligation décentralisée auprès des fournisseurs et non sur un acheteur unique (par opposition à une architecture centralisée). Ce pilier est une caractéristique distinctive du mécanisme français, qui conduit à responsabiliser directement les fournisseurs au-delà du simple financement de leur obligation de capacité. Le mécanisme est bouclé sur le réalisé côté obligation à l'aide des consommations observées mais également côté certification via la disponibilité constatée des exploitants ce qui les incite à se rendre disponible lors des périodes de pointe (par opposition à un mécanisme qui ferait le choix de rémunérer la puissance installée). Plusieurs arguments ont été avancés avant le lancement du mécanisme de capacité pour expliquer ces choix. RTE propose, à travers le retour d'expérience sur le mécanisme de capacité, d'étudier le comportement des acteurs et plus largement le fonctionnement du marché dans ce cadre règlementaire.

5.1. Le choix d'une architecture décentralisée, rebouclant sur le réalisé

Les dispositions des articles L. 335-1 et suivants du code de l'énergie établissent un dispositif d'obligation de capacité. Celui-ci prévoit que « chaque fournisseur d'électricité contribue, en fonction des caractéristiques de consommation de ses clients, en puissance et en énergie, sur le territoire métropolitain continental, à la sécurité d'approvisionnement en électricité ». Compte tenu de l'architecture décentralisée du mécanisme de capacité, les acteurs portant l'obligation de capacité ainsi créée (acteurs obligés) sont en conséquence responsables de prévoir les besoins en capacité et de se procurer les garanties de capacité nécessaires à la couverture de ce besoin.

Ces garanties peuvent être obtenues auprès des exploitants de capacité, sur le marché, ou bien directement au titre des moyens de production dont ils disposent pour les acteurs intégrés. Les acteurs obligés ont également la possibilité de diminuer la consommation de pointe de leur portefeuille par des offres adaptées (maîtrise de la consommation de pointe) pour réduire leur besoin en garanties de capacité.

Une fois l'année de livraison écoulée, le niveau de consommation réalisé lors des jours de période de pointe (jours PP1) et la disponibilité effective des capacités (sur les jours PP2) des exploitants de capacité sont évalués et peuvent être pénalisés dans le cas où des écarts seraient constatés entre le volume de garanties de capacité détenu pour les acteurs obligés et le niveau de disponibilité certifié par les exploitants. Le bouclage du mécanisme de capacité français sur les niveaux de disponibilité et de consommation réalisés, assorti d'un dispositif de règlement des écarts, a pour objectif d'inciter les acteurs obligés à couvrir leur obligation et aux exploitants de capacité de se certifier, au plus proche de leurs anticipations.

Le caractère décentralisé du mécanisme de capacité a été historiquement justifié par les raisons suivantes[1] :

- Incitation des consommateurs à modifier leur consommation pour limiter leur contribution à la pointe (i.e. développer des effacements implicites) et prise en compte les modifications effectives

  Un argument important avancé pour justifier une architecture décentralisé portait sur l'incitation renvoyée aux fournisseurs et aux consommateurs pour réduire leur consommation lors des périodes de plus forte tension pour le système électrique et sur le fait que les modifications qui en découlent sur la courbe de charge soient effectivement prises en compte dans le dimensionnement des besoins en capacité.

- Capacité à transmettre des incitations dans des temporalités courtes

  L'architecture décentralisée conduit à une potentielle réaction des prix du marché lors de chaque évènement venant affecter la vision de l'équilibre entre l'offre et la demande et in fine à l'émergence de capacité en réaction aux évolutions de prix. Ce choix d'architecture est ainsi théoriquement capable d'inciter à l'émergence de capacité dès qu'un aléa affecte la capacité du système électrique à respecter le critère de sécurité d'approvisionnement.

Il était aussi implicitement fait l'hypothèse que les acteurs obligés seraient capables d'effectuer des prévisions correctes de leur niveau d'obligation et de se couvrir en obligation en conséquence. Cette hypothèse est confrontée à l'observation du comportement des acteurs obligés lors des premières années de livraison (section 5.2 et 5.3).

L'aspect décentralisé et le bouclage sur le réalisé du mécanisme de capacité concerne également les exploitants de capacité qui doivent estimer au mieux leur niveau de disponibilité prévisionnelle (section 5.4) pour se certifier et minimiser les rééquilibrages qui peuvent engendrer des frais financiers (seules certaines capacités peuvent éviter d'être confrontées à leur disponibilité effective[2]).

Le besoin des acteurs obligés de couvrir leur obligation et l'incitation des exploitants de capacité à vendre permet l'émergence d'un prix de la capacité lors des enchères dont la formation dans un marché décentralisé pose question (section 5.5).

Enfin, les conséquences de la crise sanitaire sur le système électrique ont permis de tester la capacité de l'architecture à renvoyer des incitations de court terme via une gestion dynamique de la sécurité d'approvisionnement puisque sa temporalité ouvre la possibilité d'un dimensionnement fin du parc pour atteindre le critère de sécurité d'approvisionnement (section 5.6).

[1] Voir le rapport d'accompagnement de la proposition de règles (2014). Une justification supplémentaire était présentée et portait sur la cohérence avec la philosophie du marché européen de l'énergie. Cependant, cette cohérence n'était pas présentée comme apportant en soi des bénéfices

[2] Il s'agit des filières de production éolien, solaire et fil de l'eau qui peuvent effectuer une certification normative compte tenu de l'incertitude de leur source d'énergie primaire.

5.2. La somme des anticipations décentralisées du niveau d'obligation est de qualité analogue à celle d'une prévision « centralisée » pour les exercices AL 2017 et AL 2018

La pertinence d'une architecture décentralisée côté demande (pour le mécanisme de capacité au même titre que pour d'autres mécanismes de marché) repose sur le postulat selon lequel les acteurs de marché sont à même d'estimer leur niveau d'obligation puisqu'ils sont directement en contact avec les consommateurs et disposent donc d'une expertise unique sur leurs comportements et leurs habitudes de consommation.

5.2.1. Au niveau France, de manière agrégé, les anticipations décentralisées et anticipations centralisées de demande sont proches

Pour s'assurer que les acteurs obligés sont capables d'anticiper le niveau d'obligation qu'ils ont à couvrir, les règles du mécanisme de capacité prévoient que les acteurs obligés transmettent à RTE leur pré-estimation d'obligation deux mois avant le début de l'année de livraison[3].

L'ensemble des pré-estimations reçues sont sommés et publiées par RTE sur son Portail Services afin de participer au bon fonctionnement du marché de capacité en fournissant une estimation, vue des acteurs, du niveau de demande en garanties de capacité attendu.

Pour les années de livraison 2017 et 2018, le taux de complétude en nombre d'acteur obligé ayant transmis une pré-estimation d'obligation est respectivement de 61 % et de 76 %, ce qui correspond en termes de volume d'obligation à respectivement 1,3 GW et 1 GW[4] non remontée. Une fois cette correction de périmètre effectué, la somme des pré-estimations d'obligation reçues se trouve proche de l'obligation définitive (moins de 1% d'écart).

[3] Il s'agit d'une donnée déclarative et n'entraîne pas de conséquence financière en cas d'écart à l'obligation définitive. Néanmoins, RTE publie la liste des acteurs obligés n'ayant pas fait remonter de pré-estimation (stratégie de « name and shame »)

[4] Calculé à partir de l'obligation définitive des acteurs n'ayant pas fait remonter de pré-estimation

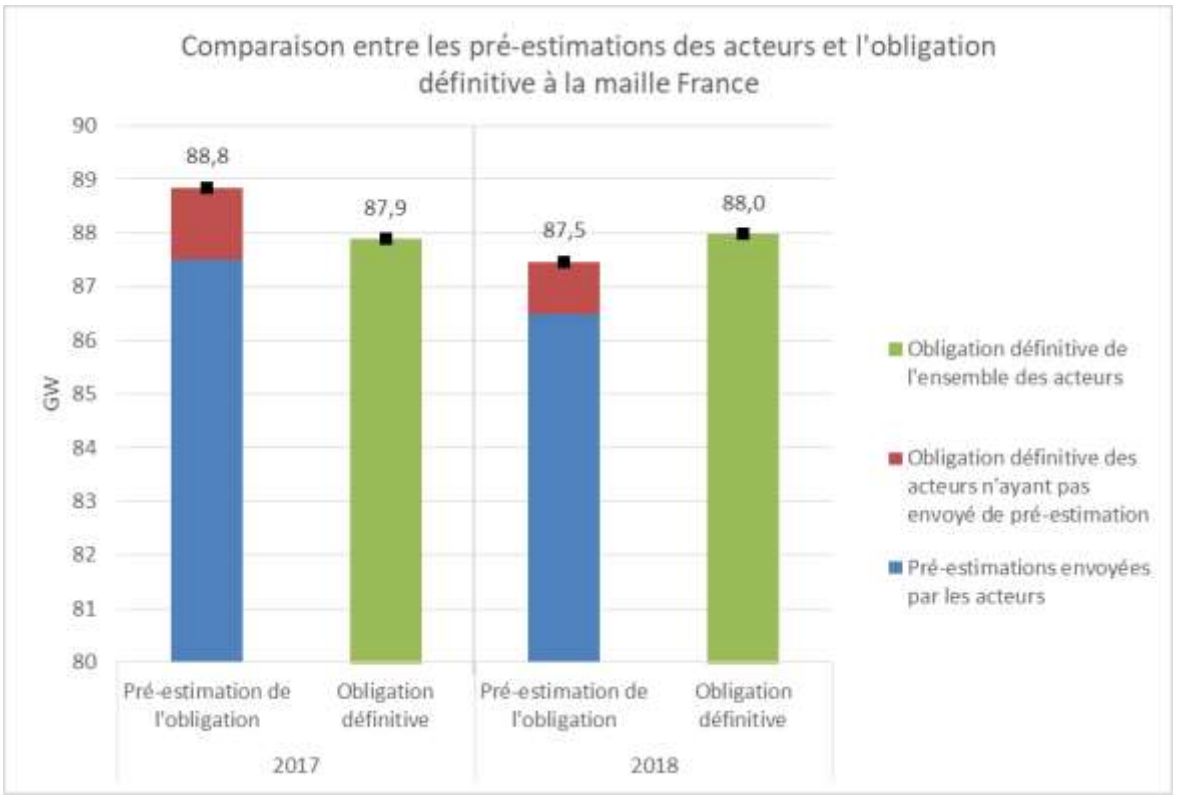


Figure 1 – Ecart entre l'anticipation décentralisée des acteurs et l'obligation définitive au niveau France

RTE réalise également un exercice centralisé de prévisions d'obligation France, dont les niveaux sont le reflet des trajectoires de consommation étudiées dans le cadre du Bilan Prévisionnel et peuvent servir de point de comparaison pour juger de la qualité de la pré-estimation décentralisée des acteurs.

Les prévisions d'obligation effectuées par RTE sont publiées à un rythme annuel et permettent ainsi d'observer l'évolution de l'anticipation centralisée plusieurs années avant le début de l'année de livraison (au contraire de la pré-estimation des acteurs qui n'est disponible que deux mois avant le début de l'année de livraison).

Il ressort de l'analyse de ces prévisions d'obligation depuis 2014 pour les années de livraison 2017 et 2018 que :

- RTE a eu une tendance à la surestimation de l'obligation plusieurs années avant le début de l'année de livraison, ce qui s'explique par stabilisation de la pointe de consommation électrique depuis quelques années en France ;
- La dernière prévision d'obligation, qui a lieu à la même échéance que la pré-estimation des acteurs obligés, montre une anticipation centralisée proche de l'anticipation décentralisée, avec un écart de l'ordre de 1% par rapport au niveau de l'obligation définitive.

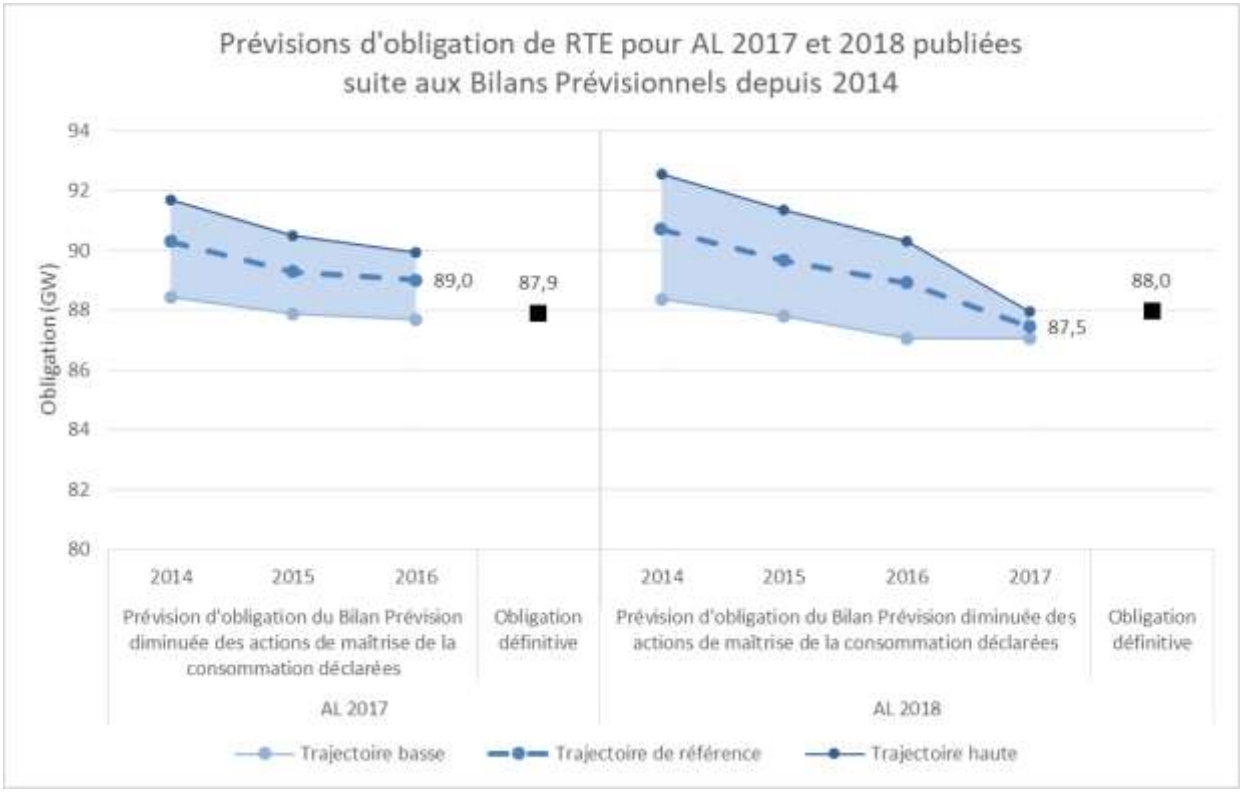


Figure 2 – Comparaison entre les prévisions d'obligation pour AL 2017 et 2018 depuis 2014 et les obligations définitives AL 2017 et 2018

En conclusion, les anticipations centralisées et décentralisés d'obligation s'avèrent proches des niveaux d'obligation définitifs observés lors des années de livraison 2017 et 2018.

5.2.2. Les petits acteurs ont plus de difficulté à anticiper leur obligation, avec un impact faible en termes de volume

Pour évaluer la capacité des acteurs obligés à estimer leur niveau d'obligation définitif, RTE a également étudié les écarts entre les pré-estimations individuelles et le niveau de l'obligation définitif par acteur. En effet, la somme des pré-estimations correcte des acteurs à la maille France peut cacher des variations importantes à la maille individuelle.

De l'analyse acteur obligé par acteur obligé, deux tendances se dégagent selon le volume du portefeuille des acteurs :

- Les acteurs obligés de plus petite taille ont eu plus de difficultés à anticiper leur niveau d'obligation puisque l'écart moyen entre la pré-estimation et l'obligation définitive pour les acteurs obligés de moins de 10 MW s'élève à près de 30%[5]. A l'inverse, les acteurs obligés de taille importante (> 1 GW) font preuve d'une meilleure capacité d'anticipation (moins de 5% d'écart) ;
- En termes de volume d'obligation, ce sont les acteurs obligés de plus grande taille qui expliquent la majorité de l'écart observé au niveau France.

[5] Les écarts en proportion par acteur obligé ont été plafonnés à 100% pour limiter l'influence des points extrêmes sur les résultats.

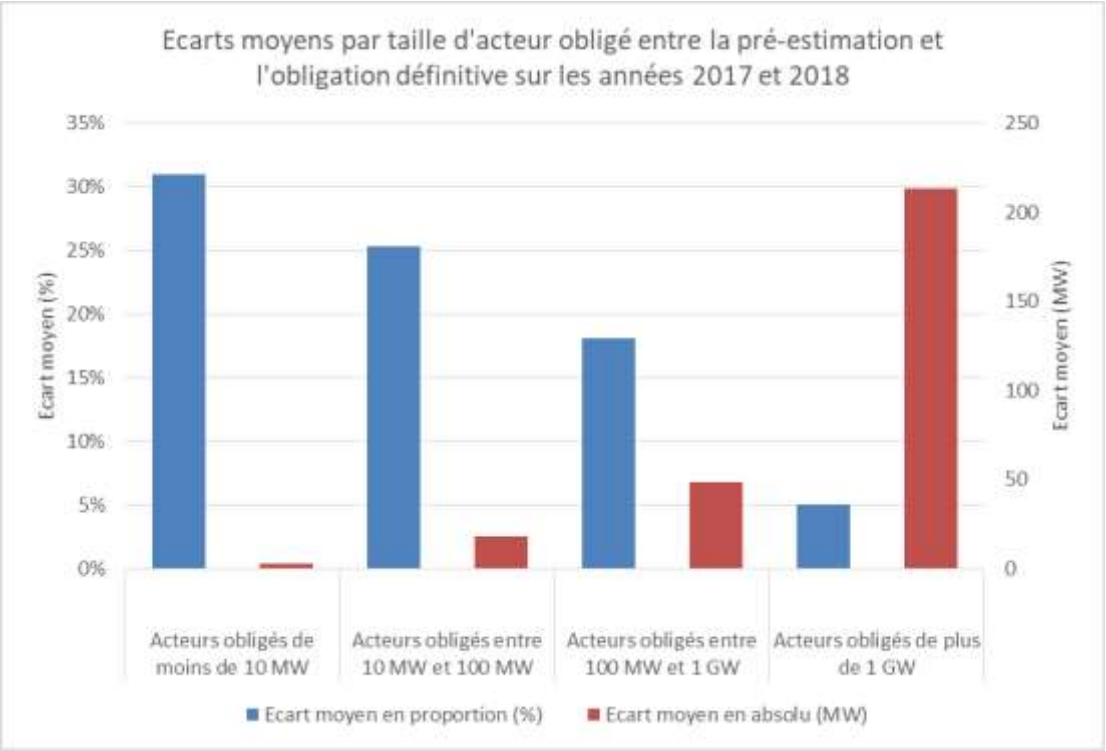


Figure 3 – Ecarts moyens par taille d'acteur obligé entre la pré-estimation et l'obligation définitive (en proportion et en absolu) pour les années de livraison 2017 et 2018

Les écarts observés à la maille acteur entre leur pré-estimation d'obligation et leur obligation définitive ne sont pas uniquement le résultat de la qualité d'anticipation des acteurs. En effet, des paramètres exogènes à l'exercice de prévision peuvent faire varier le niveau d'obligation :

- L'incertitude sur le placement des jours PP1 comme détaillée dans la section 6.4 du chapitre 6 ;
- Les modifications de portefeuille qui, en proportion, peuvent avoir un impact important sur les acteurs obligés de petite taille mais qui ont tendance à foisonner au niveau global.

5.3. Près de 5% du besoin en garanties de capacité à la maille France n'est pas couvert avant l'année de livraison

Dans l'architecture du mécanisme de capacité français, les acteurs obligés sont incités à couvrir leur besoin en capacité dans la mesure où ils sont pénalisés financièrement en cas d'écart entre le nombre de garanties de capacité détenu et le niveau d'obligation définitif lors des règlements financiers ayant lieu en mars AL+3.

La liberté laissée aux acteurs obligés en ce qui concerne la couverture de leur besoin tant en terme de temporalité (de l'ouverture des années de livraison jusqu'au débouclage en AL+3) que de leviers à leur main (achat de garanties de capacité sur le marché ou développement d'effacement implicite) a pour objectif de les responsabiliser et peut avoir pour effet de minimiser le coût d'approvisionnement en garanties de capacité pour le consommateur.

5.3.1. Une partie de l'obligation est couverte après le début de l'année de livraison

Le calcul de la couverture de l'obligation est le résultat de la comparaison entre les garanties de capacité détenues par les comptes acteurs obligés à une date donnée et le niveau d'obligation définitif. Afin de ne pas sous-estimer le niveau de couverture des acteurs obligés, à une date donnée, il est nécessaire de procéder à plusieurs ajustements d'ordre temporel :

- Les garanties de capacités obtenues via l'ARENH sont considérées comme obtenues dès le résultat du guichet ARENH de novembre (en pratique, il existe un délai de traitement entre le résultat du guichet ARENH et la réception des garanties de capacité sur les comptes des acteurs obligés).
- Les garanties de capacité obtenues au titre du tarif de cession pour les TRV des ELD[6] sont considérées comme obtenues dès le début de l'année de livraison (en pratique, les garanties de capacité liée au tarif de cession peuvent être transférées après l'année de livraison).
- Les acteurs intégrés peuvent détenir des garanties de capacité sur d'autres comptes que leur compte acteur obligé tout en les réservant pour la couverture de leur obligation. Lorsque le flux futur de garanties de capacité obtenues par un compte acteur obligé via un de ses autres comptes est plus important que le flux futur de garanties de capacité cédées par ce compte acteur obligé, l'acteur est considéré comme intégré et ce flux positif net est comptabilisé dans la couverture de l'acteur obligé[7].

La majorité de l'obligation est couverte en amont de l'année de livraison, ce qui s'explique par les cessions internes réalisés par les acteurs intégrés (environ 55%), par l'achat de garanties de capacité sur le marché ou en gré-à-gré (environ 30%) et, dans une moindre mesure, des garanties de capacité obtenues via l'ARENH (de l'ordre de 10%[8]).

Malgré la capacité des acteurs à anticiper leur niveau d'obligation définitif, vérifié en section 5.2, les acteurs obligés n'ont pas couvert la totalité de leur pré-estimation d'obligation avant le début de l'année de livraison. En effet, au 1er janvier AL, seul 95% du besoin était couvert soit de l'ordre de 5 GW d'obligations obtenues par les acteurs obligés au cours de l'année de livraison.

Le choix d'une couverture tardive de la part des acteurs obligés peut s'expliquer par des questions de trésorerie qui peuvent pousser les acteurs obligés à rapprocher leurs achats de garantie de capacité de la période de facturation de leurs consommateurs. Cela peut également provenir de l'incertitude portant sur l'évolution du besoin en garanties de capacité des acteurs obligés entre l'ouverture de l'année de livraison (début des possibilités d'échange) et la fin de l'année de livraison (consommation réalisée connue), qui provient (i) des changements du portefeuille client, (ii) du placement des jours PP1 au cours de l'année et (iii) de la méthode de calcul du gradient (voir chapitre 6). Ils profitent alors d'un marché excédentaire en garanties de capacité pour attendre une meilleure révélation de l'information concernant leur niveau d'obligation en conservant l'option de la mise en place d'effacement implicite et éviter *in fine* la surcouverture de leur obligation.

[6] Entreprise locale de distribution

[7] Si le flux futur d'un compte acteur obligé est négatif (i.e. position net vendeur), son niveau de couverture est défini uniquement sur la base des garanties de capacité détenues à une date donnée. Cette hypothèse peut conduire à surestimer la couverture des acteurs obligés utilisant leur compte AO dans une logique de *trading* (achat puis revente).

[8] Pour les années de livraison 2017 et 2018. Depuis 2019, environ 15% de l'obligation France est couverte par des garanties de capacité ARENH.

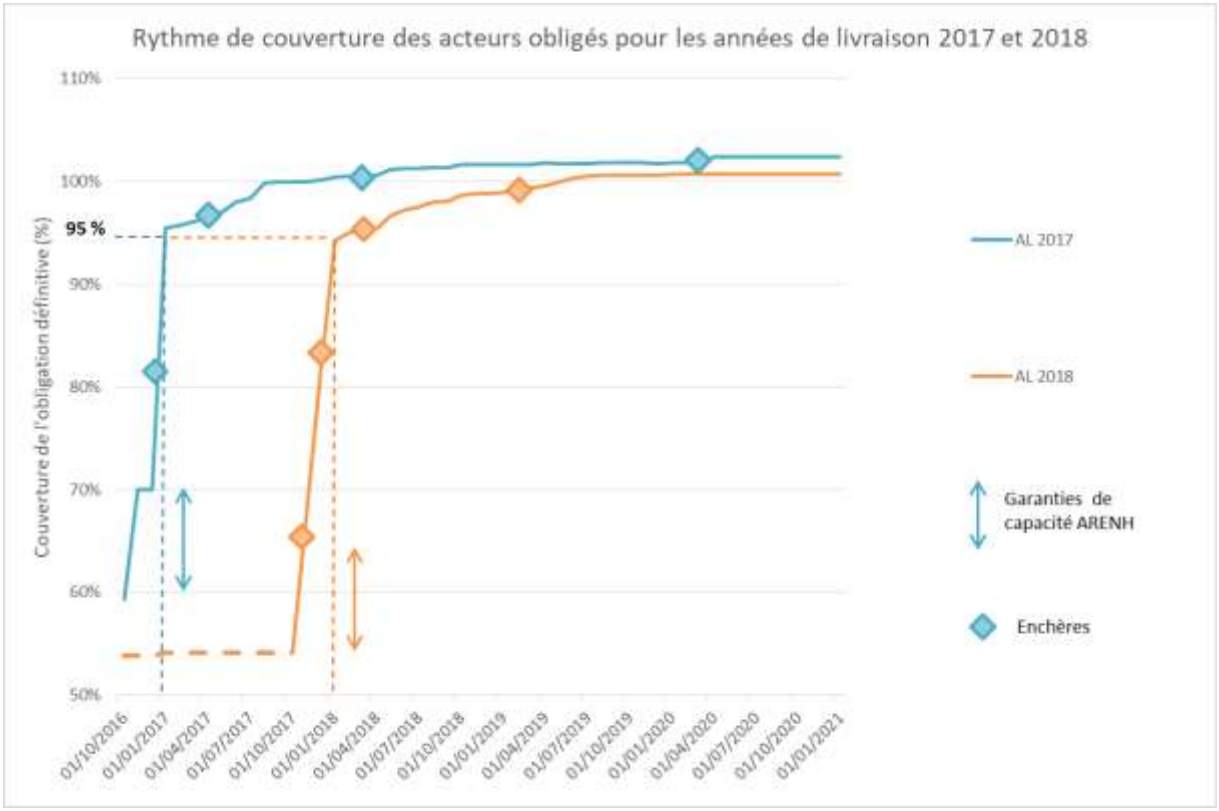


Figure 4 – Rythme de couverture des acteurs obligés pour les AL 2017 et 2018

Cependant, cette couverture tardive peut être problématique à plusieurs titres :

- Le prix qui émerge de la dernière enchère avant le début de l'année de livraison est particulièrement important puisqu'il participe à la détermination du prix de référence pour les écarts ainsi que pour les appels d'offre long terme (AOLT) et effacement (AOE). L'absence d'une partie conséquente de la demande restante en garanties de capacité à cette échéance est de nature à questionner la formation du prix lors de cette enchère de référence (à partir d'AL 2020, le prix de référence est déterminé uniquement sur la base de cette enchère). De plus, RTE constate que le nombre de garanties de capacité invendues lors de cette enchère était supérieur aux 5 GW non couverts par les acteurs obligés au 1er janvier[9], ainsi la couverture tardive des acteurs obligés est le résultat de leur stratégie d'enchère et de couverture et ne s'explique pas par un manque de liquidité sur le marché de capacité.
- La disponibilité des capacités correspondantes au besoin non couvert par les acteurs obligés (environ 5 GW) avant le 1er janvier s'est avérée nécessaire au respect du critère de sécurité d'approvisionnement. Bien que ces capacités aient pu vendre leurs garanties de capacité après le début de l'année de livraison, leurs revenus capacitaires n'ont pas été sécurisés en amont de l'année de livraison au moment où les exploitants arbitrent entre maintien et fermeture du moyen de production.
- Compte tenu de la multiplicité des enchères et la liberté donnée aux acteurs obligés concernant la temporalité de leur couverture, les offreurs de capacité peuvent être amenés à

[9] Conformément au plafond de prix sur les exercices AL 2017 et AL 2018, ces garanties de capacité ont été proposées à un prix inférieur à 20 000 €/MW.

adopter des stratégies non-conforme à la théorie économique du *missing money* (voir section 5.5 pour plus de détail).

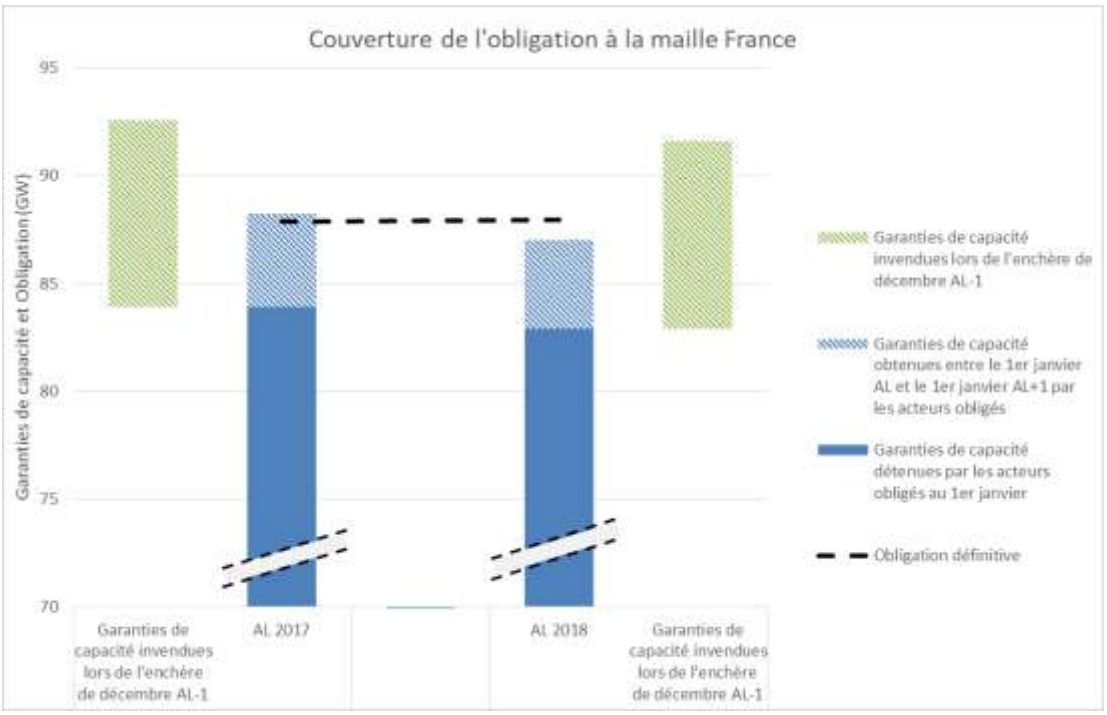


Figure 5 – Situation de la couverture des acteurs obligés au 1er janvier AL

5.3.2. En proportion les plus petits acteurs obligés sont moins bien couverts mais le volume non couvert est principalement le résultat des acteurs obligés de plus grande taille

De manière analogue aux écarts constatés entre les pré-estimations d'obligation et l'obligation définitive, RTE remarque que les plus petits acteurs obligés sont moins bien couverts que les gros acteurs obligés au début de l'année de livraison.

Pour autant, le volume non-couvert est en majeure partie (70%) le résultat de la couverture partielle des acteurs obligés de plus grande taille.

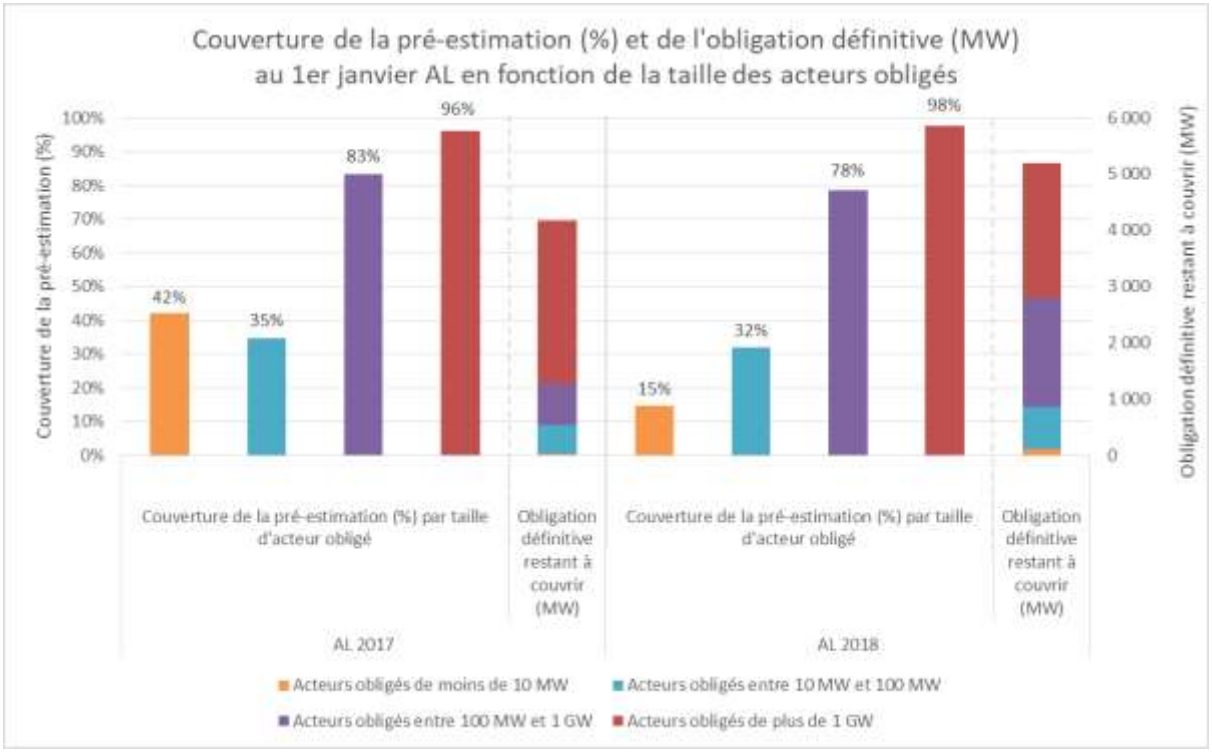


Figure 6 – Pour les acteurs ayant envoyé une pré-estimation, comparaison entre le % de couverture de la pré-estimation et le volume de couverture manquant en fonction de la taille des acteurs obligés

5.3.3. Le développement des effacements implicites est resté limité

L'une des justifications au choix d'une architecture décentralisée portait sur la responsabilité donnée aux acteurs obligés qui leur permet d'influer, au travers de leurs pratiques commerciales ou tarifaires, sur les structures de consommation de leur client. L'intérêt d'un système décentralisé est faible si les acteurs obligés sont passifs par rapport à l'obligation. Au contraire, le choix de l'architecture décentralisée doit permettre d'engager pleinement la maitrise de la pointe de consommation dans le maintien de la sécurité d'approvisionnement française en promouvant les effacements des consommateurs lors des épisodes de tension sur le système électrique.

Le caractère allocatif du mécanisme de capacité français, c'est-à-dire sa faculté à transmettre à chaque groupe de consommateur le signal économique de son impact sur la sécurité d'approvisionnement (vérifié en 3.2.3), a pour effet d'inciter à la réduction de la pointe de consommation. Ces réductions de consommation, appelés effacements implicites, désignent les capacités d'effacement indissociables de la fourniture et qui sont valorisées par les fournisseurs par une réduction de leurs coûts d'approvisionnement (à la fois en énergie et en capacité, à travers une réduction de l'obligation de capacité). Ces effacements implicites doivent, selon les l'article R335-66 du code de l'énergie, faire l'objet d'une déclaration par les fournisseurs, afin de permettre à l'ensemble des acteurs du marché de capacité d'anticiper le niveau d'obligation global France.

A ce titre, RTE a mis en place un registre de maîtrise de la consommation de pointe qui a pour objectif de recenser les effacements implicites déclarés par les acteurs obligés avant le début de l'année de livraison[10]. A quelques exceptions près, seuls les effacements indissociables de la fourniture des tarifs

[10] Article 11.5 des Règles du mécanisme de capacité

réglementés de vente (option tarifaire Tempo et EJP) ont donné lieu à des déclarations de volume d'effacement implicite, de l'ordre de 700 MW entre 2017 et 2019.

RTE a procédé à une analyse de la consommation lors des jours PP1 afin d'y déceler de possibles effacements de consommation de pointe non déclarés au sein du registre de maîtrise de la consommation de pointe. RTE a remarqué que les niveaux de consommation observés sur les plages horaires PP1 sont statistiquement inférieurs aux prévisions de consommation corrigées des erreurs de prévision de température ce qui suggère que des effacements indissociables de la fourniture supplémentaires auraient été mobilisés. Cet effet est chiffré à environ 300 MW, mais reste sujet à caution étant donné l'incertitude concernant son évaluation (+/- 100 MW) et l'incitation du prix de l'énergie concomitante à celle du prix de la capacité. Le volume d'effacement implicite ne s'est donc pas développé massivement ce qui est confirmé par l'absence de développement d'offre de marché sur le signal TEMPO.

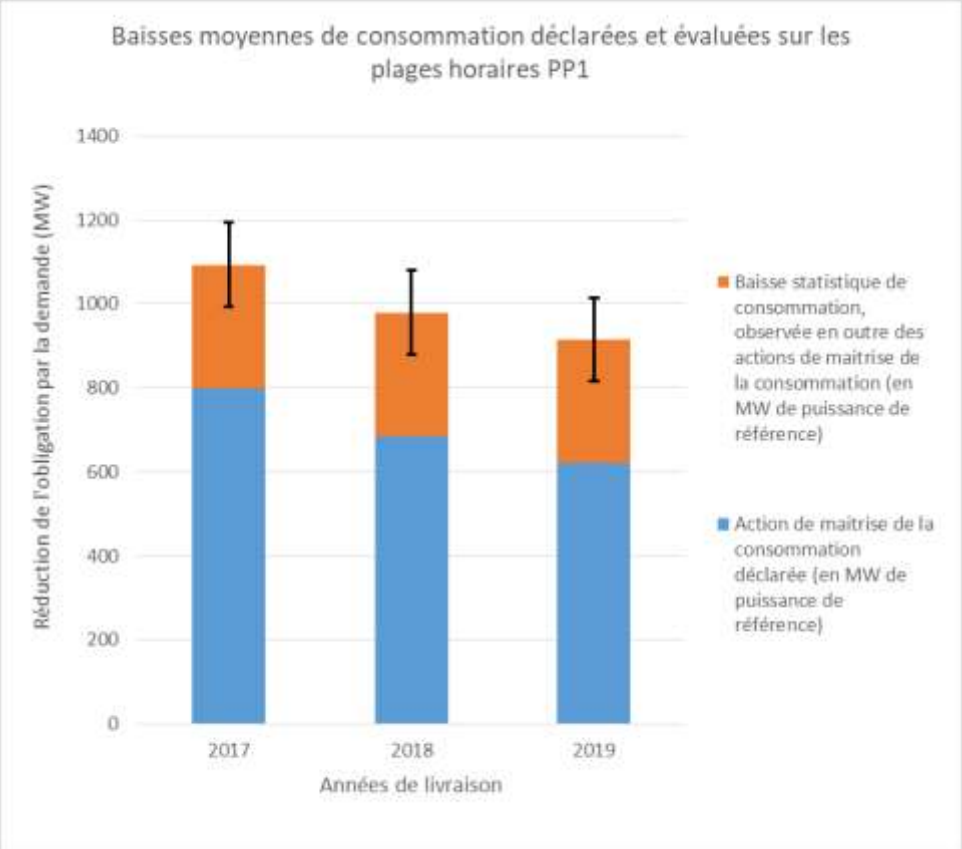


Figure 7 – Evolution des volumes d'effacement implicites au cours des trois premières années de livraison

5.4. L'anticipation décentralisée de la disponibilité des exploitants de capacité a été largement surestimée sur les premières années de livraison

Dans l'architecture décentralisée et rebouclant sur le réalisé du mécanisme de capacité français, les exploitants de capacité sont responsables de l'estimation de leur niveau de disponibilité prévisionnelle et de la vente de leurs garanties de capacité. Le cadre règlementaire portant sur les frais de rééquilibrages et le règlement des écarts (entre le niveau de capacité certifié et la disponibilité constatée) a pour objectif de les inciter à estimer le plus finement possible et au plus tôt leur disponibilité prévisionnelle.

5.4.1. L'aptitude des exploitants de capacité à estimer leur disponibilité en amont de l'année de livraison est remise en question à la lumière des premières années de livraison

La capacité des acteurs obligés à anticiper leur niveau de disponibilité (et donc leur niveau de capacité effectif – NCE) peut s'évaluer à partir des deux indicateurs, mis à disposition par RTE sur son Portail Services, que sont le NCC – Niveau de capacité certifié – et le NCC évolué – Niveau de capacité certifié évolué :

- Le NCC correspond au niveau de disponibilité contractualisé par les exploitants auprès de RTE, il donne lieu au versement de garanties de capacité valorisables sur le marché. L'évolution du NCC au cours du temps est le résultat des rééquilibrages pouvant exposer les acteurs à des frais financiers, tout comme l'écart qui serait constaté en AL+3 entre le dernier NCC et le NCE ;
- Le NCC évolué est une donnée déclarative et non contractuelle, qui correspond à la meilleure vision à date de leur disponibilité prévisionnelle (la dernière déclaration des exploitants ou dernier contrat signé en l'absence de déclaration).

Ces deux indicateurs peuvent évoluer chaque jour et fournissent ainsi une vision dynamique de la disponibilité prévisionnelle du parc lors des périodes de tension (les jours PP2). L'écart entre le NCC et le NCC évolué peut relever du choix d'un acteur de ne pas se rééquilibrer au plus proche de sa meilleure prévision de disponibilité [11] ou du délai de traitement des demandes de rééquilibrages plus chronophage que les demandes d'évolution de paramètre[12]. L'ensemble des rééquilibrages présentés par la suite sont comptabilisés à la date de demande des acteurs afin de ne pas être affectés par les délais d'instruction inhérents à la signature d'un contrat.

[11] Par exemple, compte tenu des frais financiers de rééquilibrage (voir chapitre 8)
[12] Re-déclaration de la disponibilité qui ne nécessite pas la signature d'un contrat de certification et qui modifie uniquement le NCC évolué (obligatoire pour toute variation d'au moins 100 MW du niveau de capacité certifié)

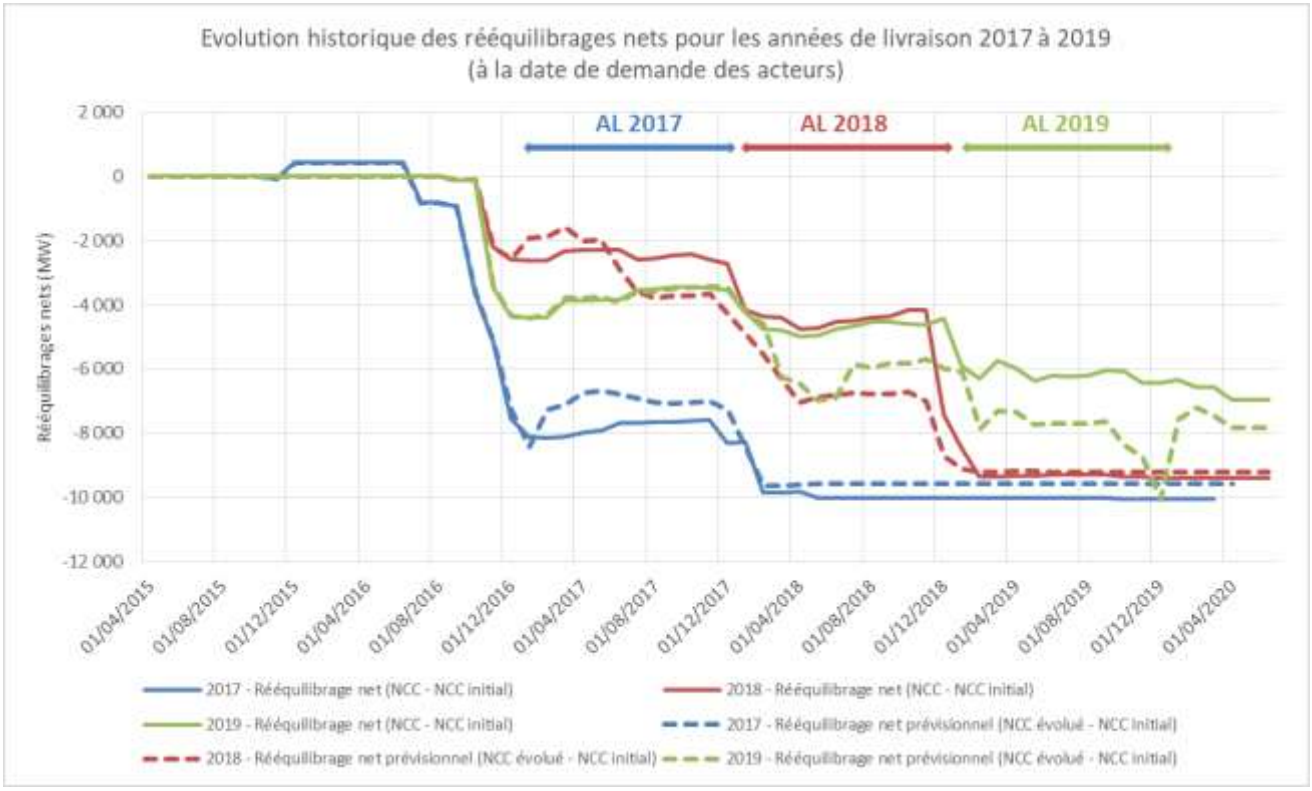


Figure 8 – Evolution du volume de rééquilibrage net pour AL 2017 à 2019

En comparant les niveaux de disponibilités contractualisés et les niveaux de disponibilité prévisionnelle à la certification initiale des exploitants, il est possible de connaître l'évolution historique du volume de rééquilibrage net et anticipé par les acteurs dont plusieurs constats peuvent être tirés :

- Les exploitants de capacité ont largement surestimé leur niveau de disponibilité lors de leur certification initiale puisque les rééquilibrages successifs ont conduit à un niveau de capacité certifié final jusqu'à 10 GW (AL 2017) plus faible que le niveau de capacité certifié initialement. Ce biais à l'optimisme a donné lieu à des rééquilibrages massifs quelques mois avant le début des années de livraison, en particulier près de 30% du volume de rééquilibrage a lieu entre l'enchère de référence[13] et le 1er janvier AL ;

- Le niveau de disponibilité prévisionnelle[14] s'est trouvé inférieur (jusqu'à 3 GW au cours d'année de livraison 2018) au niveau de capacité certifié la majeur partie du temps sur les trois premières années de fonctionnement du mécanisme de capacité. Compte tenu de l'orientation des rééquilibrages à la baisse, le NCC évolué a pu constituer une anticipation des évolutions du NCC, ce qui prouve que (i) les acteurs avaient conscience de leur biais à l'optimisme et (ii) n'ont pas réalisé, au plus tôt, de rééquilibrages à leur meilleure prévision de disponibilité (voir chapitre 8) qui se seraient trouvés justifiés a posteriori.

- Il semble exister de premiers effets d'apprentissage puisque le biais à l'optimisme paraît se réduire pour l'année de livraison 2019, pour laquelle le volume de rééquilibrage net est inférieur aux AL 2017 et 2018 et que la plupart de ces rééquilibrages ont pu être effectués plusieurs années en amont de l'année de livraison. De plus, c'est la première année de livraison qui n'a pas donné lieu à des rééquilibrages de plusieurs GW après la fin de l'année de livraison.

[13] Dernière enchère avant le début de l'année de livraison
[14] Appelé également NCC évolué

L'ensemble des filières sont concernés par le biais à l'optimisme détecté à la maille nationale. La filière nucléaire porte plus de 50% des volumes rééquilibrés nets en cohérence avec sa contribution à la sécurité d'approvisionnement française. Seules les entités relatives aux filières effacements et EnR font apparaître des rééquilibrages orientés à la hausse mais cela s'explique essentiellement par leur manière de procéder à l'ajout de site après la certification initiale[15].

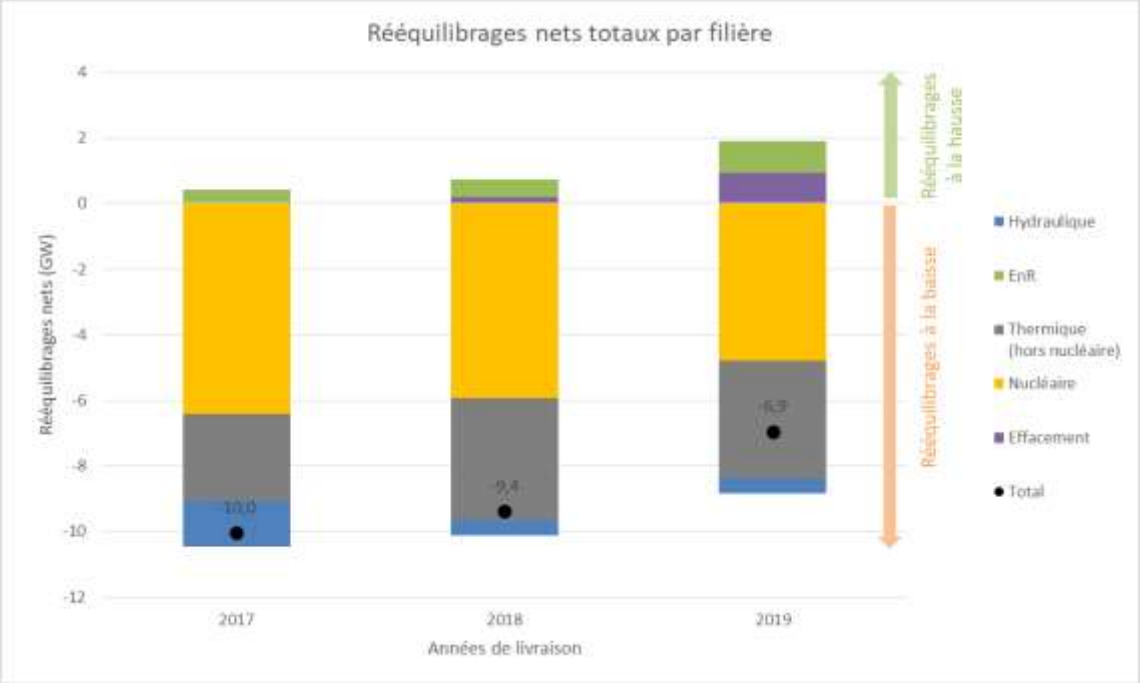


Figure 9 – Total des volumes de rééquilibrages nets décomposés par filière

5.4.2. La déclinaison des anticipations du Bilan Prévisionnel au mécanisme de capacité montre des estimations proches du niveau de disponibilité constaté

RTE réalise, chaque année, des études d'adéquation entre l'offre et la demande d'électricité sur la base des informations disponibles, des retours des acteurs à la consultation publique ainsi qu'à une représentation probabiliste des aléas affectant le système électrique (météo, disponibilité du parc de production, ...).

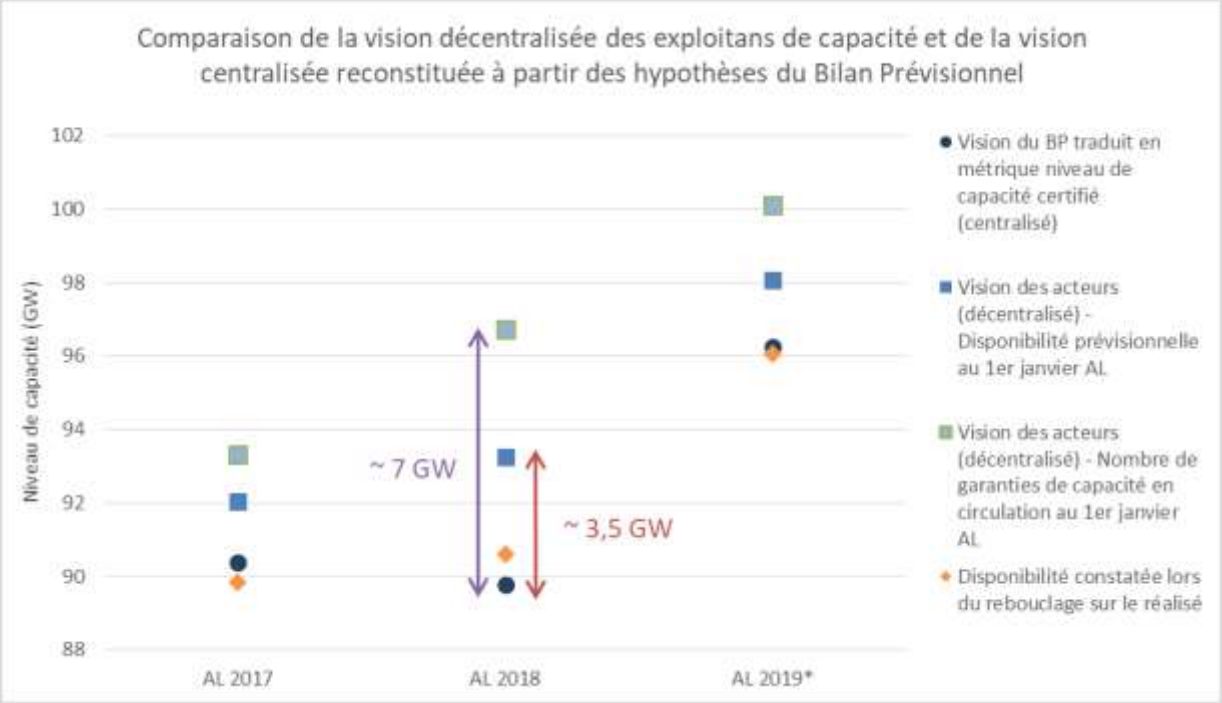
RTE a traduit les hypothèses considérées dans les Bilans prévisionnels concernant la disponibilité des capacités dans la métrique « niveau de capacité certifiée » (i.e. en appliquant les principes de calculs du niveau de certification – calcul de l'espérance de disponibilité sur la base de tirages probabilistes de jours PP2, application des coefficients « filières » …) afin de comparer les anticipations décentralisées de disponibilités des acteurs à celles centralisées du Bilan Prévisionnel.

En toute rigueur, seule la disponibilité affichée sur le mécanisme de capacité en amont de l'année livraison est comparable à celle du Bilan Prévisionnel puisqu'elles renvoient toutes deux à un même niveau d'information sur l'année de livraison à venir. Une fois l'année de livraison débutée et le

[15] En effet, certains opérateurs d'effacements procèdent dans un premier temps à la création d'une entité de certification avec quelques sites, puis ajoutent jusqu'au 31 octobre AL + 1 les sites qu'ils sont parvenus à contractualiser. Pour les capacités sous obligation d'achat, les certifications sont également effectuées sous la forme de rééquilibrage tous les deux mois, au fur à mesure de la mise en service des sites sous obligation d'achat.

premier jour PP2 tiré, les exploitants acquiert une information sur leur niveau de disponibilité qui est entièrement déterminé une fois que plus aucun jour PP2 ne peut être tiré.

La comparaison de la vision centralisée du Bilan Prévisionnel avec le niveau de disponibilité contractualisée (NCC) et de disponibilité prévisionnelle (NCC évolué) attestent de l'optimisme des acteurs quant à leur disponibilité. En effet, la disponibilité anticipée par les acteurs est nettement supérieure, pour l'ensemble des années de livraison, à la vision du Bilan Prévisionnel (p.e. de l'ordre de 3,5 GW pour 2018, vu du 1er janvier 2018 et jusqu'à 7 GW par rapport au volume de garanties de capacité en circulation à cette même date).



* Prise en compte explicite des interconnexions à partir de l'année de livraison 2019

Figure 10 – Comparaison de la vision décentralisée des exploitants de capacité avec la vision centralisée reconstituée à partir du Bilan Prévisionnel

L'analyse de la disponibilité constatée permet d'identifier que l'hypothèse de RTE établie dans le Bilan prévisionnel s'est avérée plus réaliste que l'anticipation des exploitants de capacité, reflétée dans leurs certifications, qui a été sur-évaluée systématiquement d'au moins 2 GW, mais l'écart tend à se réduire depuis 2019 confirmant de possibles effets d'apprentissage de la part des exploitants.

5.5. La formation des prix semble cohérente avec les fondamentaux, mais soulève des enjeux de transparence

Le mécanisme de capacité français fonctionne sous la forme d'un marché (*market-based*) c'est-à-dire que le prix du produit « garanties de capacité » est le résultat d'un équilibre entre l'offre et la demande (et non d'un choix administratif). Dans un marché de capacité, le prix révèle la tension sur l'équilibre offre-demande mais aussi les imperfections du marché de l'énergie (théorie du *missing-money*). Côté offre, il permet de monétiser les externalités positives sur la sécurité d'approvisionnement engendrées par la disponibilité des capacités tandis que, côté demande, il alloue aux consommateurs, les coûts

engendrés par leur consommation en période de pointe. Finalement, le prix de la capacité agit comme un signal à l'attention de l'ensemble des acteurs pour leur permettre d'agir en cohérence dans l'objectif d'atteindre du critère de sécurité d'approvisionnement.

Une formation adéquate du prix de la capacité est donc cruciale à l'atteinte des objectifs du mécanisme de capacité.

Tant la propension des acteurs obligés à ne pas couvrir totalement leur obligation de capacité avant le début de l'année de livraison que le biais à l'optimisme des exploitants de capacité quant à leur disponibilité prévisionnelle questionnent légitimement sur le niveau du prix de la capacité. RTE propose des analyses qui se concentrent sur les prix résultant des enchères organisées afin d'objectiver la formation du prix de la capacité.

5.5.1. La formation du prix semble cohérente avec les fondamentaux théoriques

Les acteurs obligés et les exploitants peuvent échanger des garanties de capacité en bilatéral, sur le marché de gré-à-gré ou encore à travers le marché organisé sous forme de sessions de marché organisés qui en régime nominal seront au moins au nombre de 15 pour chaque année de livraison[16].

Les prix formés lors des enchères font référence car ils définissent :

- le Prix de Référence Marché (PRM) – moyenne arithmétique simple des prix révélés par les enchères précédant l'année de livraison – qui sert à la construction par empilement des tarifs régulés de vente (TRV)[17] et qui était la référence pour le règlement des écarts jusqu'à l'année de livraison 2019 ;
- le Prix du Règlement des Écarts (PREC)[18] – prix qui résulte de la dernière session de marché organisée avant le démarrage de l'année de livraison – qui est la référence de prix utilisée pour les règlements financiers à partir de l'année de livraison 2020;

  et sont également utilisés comme référence de prix par de nombreux acteurs lors de leurs échanges en gré-à-gré.

Dans les conditions de la concurrence pure et parfaite[19] et d'une enchère unique engageant l'ensemble de la demande des acteurs obligés, correspondant à la contribution de leurs consommateurs au risque de défaillance, et de l'offre résultant de l'engagement des exploitants quant à la disponibilité de leur moyens, la théorie économique fournit un cadre d'analyse de la formation du prix sur le marché de capacité via le concept de « *missing money* ».

[16] Au 01/04/2021, ce sont tenues : 4 enchères pour AL 2017, 5 enchères pour AL 2018, 9 enchères pour AL 2019, 9 enchères pour AL 2020, 6 enchères pour AL 2021 et 5 enchères pour AL 2022

[17] En théorie, les TRV ne sont pas construits à partir du PRM mais à partir de la moyenne arithmétique du prix des enchères constatés sur les deux années précédant l'année de livraison. En pratique, il ne s'est jamais tenu d'enchère plus de deux ans avant une année de livraison ainsi la capacité se trouve, en pratique, valorisée au PRM dans les TRV.

[18] A partir de l'année de livraison 2020, le PREC remplace le PRM comme référence de prix pour le règlement des écarts

[19] Sous les hypothèses d'atomicité des acteurs, d'homogénéité des biens échangés, de libre entrée et sortie du marché ainsi que l'absence d'asymétrie d'information et de toute forme d'externalité.

Sur le marché de l'énergie, les producteurs offrent leur production à leur coût marginal ; ainsi dès lors que le prix de marché s'élève au-dessus de ce coût, l'exploitant est incité à produire. Le prix du marché de l'énergie se trouve ainsi fixé au coût variable de fonctionnement du dernier moyen appelé et les producteurs couvrent ainsi assurément leur coût variable de production et reçoivent une rente infra-marginale (lorsque le moyen de production est activé mais qu'un autre moyen fait le prix de marché) pour couvrir leur coût fixe. La décision d'investir ou de maintenir une capacité s'effectue sur la base de l'espérance des revenus futurs (intégrant une aversion au risque des acteurs) sur le marché de l'énergie compte tenu de l'ensemble des coûts de fonctionnement[20].

Le mécanisme de capacité doit alors permettre aux moyens de production et d'effacement existants non-rentables sur le marché de l'énergie seul, de compléter leur revenu à la mesure de leur *« missing money »*[21] sur le marché de l'énergie. Le niveau de sécurité d'approvisionnement visé est alors atteint à l'aide du prix de la capacité ainsi formé, sur le « *missing money* » du dernier moyen nécessaire à la sécurité d'approvisionnement. Si le « *missing money* » d'une capacité est supérieur au prix du marché de capacité alors cette capacité n'était pas utile à la sécurité d'approvisionnement, et à l'inverse, toutes les capacités nécessaires à l'atteinte du niveau de sécurité d'approvisionnement visé trouveront un complément de revenu sur le marché de capacité leur permettant de rester en fonctionnement[22].

La Commission de Régulation de l'Énergie (CRE) a développé un modèle pour assurer la surveillance du marché de capacité et vérifier que le prix de la capacité se forme au niveau attendu en se basant sur ce cadre théorique du « *missing money* ». En effet, le prix théorique est reconstitué comme l'intersection entre une demande à tout prix (prix administré) pour tout le volume de prévision d'obligation publié par RTE et une offre correspondant à l'empilement croissant des « *missing money* » des différentes capacités du système électrique français selon leur niveau de disponibilité anticipé issu du registre public des capacités certifiées.

Les résultats de ces analyses, publiés dans ses rapport de surveillance des marchés de gros 2018 et 2019 amènent la CRE à conclure que *« l'équilibre global du système se trouve, pour les années considérées (AL 2017 à 2019), au niveau des « missing money » des centrales thermiques à flamme »* et que *« les simulations présentées montrent que les prix issus des différentes enchères ayant eu lieu pour les années de livraison 2017 et 2018 semblent compatibles avec les fondamentaux du mécanisme de capacité »*.

[20] La décision d'investir ou de maintenir une capacité ne se fait pas uniquement sur des critères économiques, d'autres paramètres (politiques, sociaux, …) pouvant également entrer en ligne de compte.

[21] Le *missing money* est le résultat de la différence entre les coûts fixes nécessaires au fonctionnement du moyen de production (i.e. évitables par une fermeture) et les revenus apportés par les marchés de l'énergie. Il convient également d'en déduire les coûts de mise sous cocon ou de fermeture qui seraient engagés dans l'hypothèse où l'acteur choisirait de mettre fin à l'exploitation de l'actif. A noter également que le *missing money* peut évoluer dans le temps dans la mesure où les conditions de marché sont variables et que les coûts évitables diminuent en se rapprochant du temps réel.

[22] Pour les nouvelles capacités, les coûts d'investissement sont des coûts évitables qui entrent donc en jeu dans la détermination du prix d'offre.

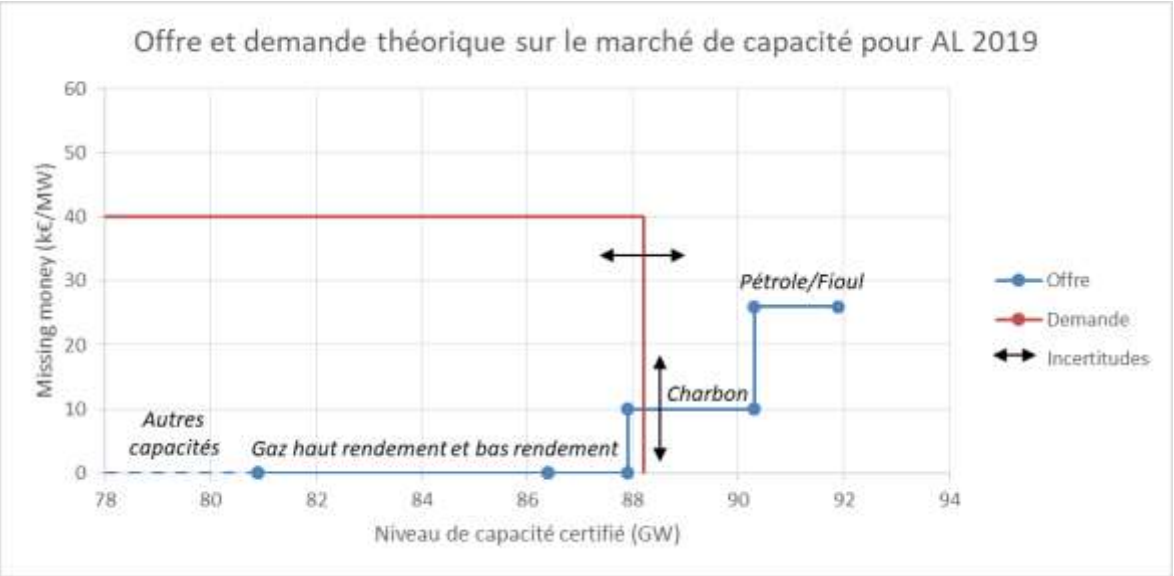


Figure 11 – Illustration de l'offre et demande théorique sur le marché de capacité pour AL 2019 (à partir de l'analyse de la CRE dans son rapport de surveillance sur les marchés de gros de 2018)

Cependant, la CRE note également « *une forte sensibilité du prix d'équilibre aux différentes hypothèses, notamment de coûts fixes des capacités de production et de niveau de la demande* » ce qui « *ne permet pas de statuer clairement quant à la capacité marginale* ».

5.5.2. Une évolution du niveau de prix cohérente avec l'évolution des marges observées lors des enchères portant sur les années de livraison 2020 et 2021

La crise sanitaire a conduit à un allongement des arrêts pour maintenance des centrales nucléaires ayant pour conséquence une moindre disponibilité du parc nucléaire pour l'hiver 2020-21 ce qui a conduit RTE à placer l'hiver sous « vigilance particulière » dès le mois de juin 2020. La situation de tension prévisionnelle sur l'équilibre offre-demande a entraîné un effet logique sur l'équilibre du marché de capacité pour les années de livraison 2020 et 2021, réduisant les excédents en certificats de capacité et faisant apparaître un déficit pour l'année 2020.

La différence entre la disponibilité prévisionnelle déclarée par les exploitants de capacité (indicateur NCC évolué) et la demande à venir pour couvrir l'obligation des acteurs obligés (prévision d'obligation France de RTE) constitue une « marge de capacité » reflétant la tension sur l'équilibre offre-demande qui est au fondement du prix de la capacité. En effet, plus la disponibilité anticipée du parc est faible, plus l'offre de capacité en sera diminuée ce qui conduira à une situation de rareté de la garantie de capacité et donc à une augmentation du prix ; à l'inverse une consommation à la pointe en berne induira une faible demande en garantie de capacité à même de faire baisser le prix. Il est donc attendu que le prix de la capacité et la marge de capacité soient reliés par une relation décroissante.

La forte volatilité du prix de la capacité et de la marge de capacité pour les années de livraison 2020 et 2021 a permis de mettre en exergue la relation attendue – figure 12 – venant confirmer la cohérence de l'évolution du prix de la capacité avec les fondamentaux.

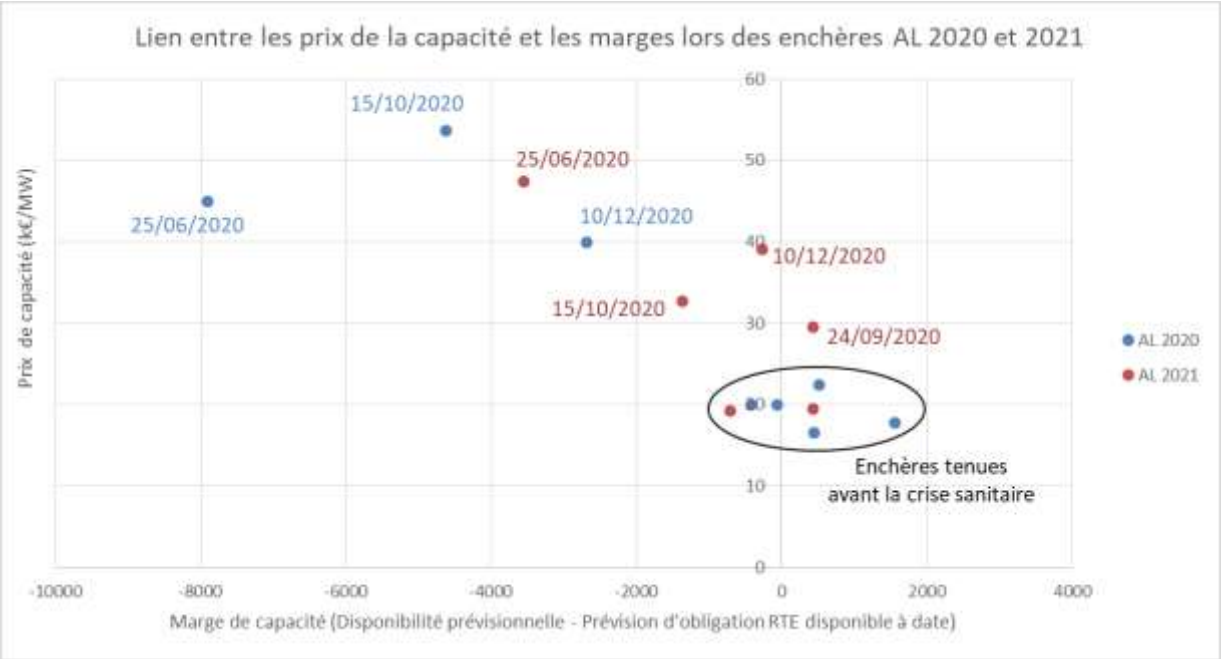


Figure 12 – Lien entre les prix de la capacité et les marges de capacité lors des enchères

Il convient néanmoins de noter que cette analyse pour expliquer la formation des prix de la capacité présente plusieurs limites dont les principales sont :

- lorsqu'une enchère a lieu avant le début de l'année de livraison, une partie du parc de capacité peut ne pas encore être certifié, comme les effacements qui peuvent effectuer une demande de certification sans frais jusqu'au 31 octobre AL-1.
- La marge de capacité ne constitue pas le seul facteur explicatif du niveau du prix de la capacité. En effet, la formation du prix de la capacité dépend également du « *missing money* » de la capacité marginale, qui évolue en fonction de ses coûts fixes et de ses revenus sur le marché de l'énergie.
- Compte tenu des différents paliers de « *missing money* », la relation décroissante entre le prix de la capacité et la marge de capacité n'est pas nécessairement linéaire.

5.5.3. Des écarts à la théorie économique du missing money qui crée un enjeu de transparence

Dans les mêmes rapports de surveillance du marché de gros (2018 et 2019) qui ont conduit à conclure à la compatibilité des fondamentaux du mécanisme de capacité avec la formation du prix, la CRE a exprimé sa préoccupation par rapport au fonctionnement observé des enchères qui ne se caractérisent pas par des prix d'offre reflétant le niveau de missing money par technologie, ni par un niveau de demande à tout prix (figure 13), comme cela était anticipé en théorie. Cette préoccupation par rapport à la transparence de la formation du prix lors des enchères a été largement partagée par les acteurs de la concertation qui font le constat qu'un large palier d'offre influence fortement la formation du prix lors des sessions de marché organisé. Ainsi, sur la base de ce fonctionnement empirique, la CRE explique que « *l'architecture du mécanisme ne permet pas la rencontre efficace de l'offre et de la demande et conduit certains acteurs à ne pas offrir leurs garanties de capacité au niveau du « missing money » de leurs capacités.* »

En théorie, les capacités dont les revenus sont assurés par le seul marché de l'énergie devraient être amenées à offrir leurs garanties de capacité à un prix nul. En pratique, un tel comportement n'est pas observé sur le marché de capacité, certains offreurs de capacité faisant le choix de proposer leur garantie de capacité à un prix de réserve.

Ce comportement peut s'expliquer par l'incertitude pesant sur le niveau de demande lors des enchères organisées compte tenu (i) de la temporalité du mécanisme autorisant les acteurs obligés à acheter des garanties de capacité sur plusieurs années entre l'ouverture de l'année de livraison et les règlements financiers en AL+3, ainsi que (ii) de la multiplicité des enchères et (iii) de la possibilité d'effectuer des échanges en gré-à-gré, qui ne permettent pas la rencontre de toute l'offre et toute la demande lors d'une enchère unique. A l'opposé de la liberté donnée aux acteurs obligés pour couvrir leur obligation, les exploitants de capacité pouvant se trouver en position dominante[23] sont contraints d'offrir une partie de leur garantie de capacité à chaque enchère et ainsi être amenés à mettre un prix de réserve.

Près de la moitié des enchères organisées se sont conclues à un prix au niveau du plateau d'offre ce qui crée un fort enjeu de lisibilité sur la formation du prix.

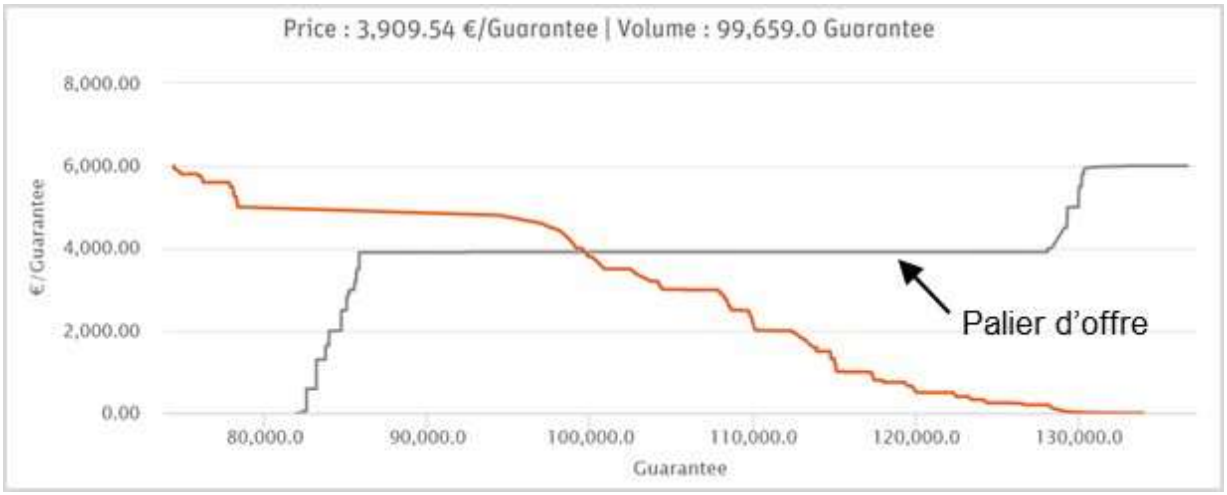


Figure 13 – Enchère de décembre 2020 pour AL 2021 (Source : EPEX SPOT)

5.6. Le prix de la capacité renvoie les incitations économiques adéquates en cas de difficulté sur la sécurité d'approvisionnement : exemple de l'hiver 2020-21

Le caractère décentralisé du mécanisme de capacité français est également lié à sa temporalité qui assure une grande adaptabilité du mécanisme ayant pour objectif d'en minimiser le coût pour le consommateur. En effet, la possibilité d'échanger des garanties de capacité jusqu'au dernier moment permet de faire émerger un prix de la capacité cohérent avec les aléas intervenant à la fois côté offre (évolution de la disponibilité, émergence de nouvelles capacités) et côté demande (contexte macroéconomique, structure de la consommation) à différentes échéances. Ainsi, le mécanisme décentralisé français offre la possibilité aux acteurs de marché de gérer ces risques sur le long comme

[23] Les exploitants de plus de 3 GW de capacités certifiées doivent offrir un niveau minimum de garanties de capacité sur chacune des enchères.

sur le court terme : en théorie, cela écarte à la fois le risque de surdimensionnement du système électrique et le risque de sous-investissement (incitation par le prix de la capacité).

La propagation de la Covid-19 sur le territoire français en 2020 a provoqué une crise sanitaire sans précédent qui n'a pas été sans impact sur le système électrique français puisque cela a entraîné une baisse de la consommation (de l'ordre de 15% de baisse au plus fort de la crise[24]) et diminution de la disponibilité anticipée du parc nucléaire pour l'hiver 2020-21. La modification profonde de ces fondamentaux a conduit à une réaction des acteurs sur le marché de capacité permise par le choix d'architecture du mécanisme français.

5.6.1. Suite à la diminution de la disponibilité du parc nucléaire en conséquence des modifications de planning compte tenu de la crise sanitaire, le prix de la capacité a augmenté …

La baisse de la disponibilité de la filière nucléaire s'est matérialisée par une baisse de plusieurs GW de l'indicateur « NCC évolué » au mois de mai 2020 (baisse de l'ordre de 7 GW pour AL 2020 et 3,5 GW pour AL 2021). RTE a évalué et communiqué dès le mois de juin 2020[25] sur le diagnostic de sécurité d'approvisionnement de l'hiver à venir qui a conduit à le placer sous vigilance particulière. La brusque modification de la situation de sécurité d'approvisionnement s'est traduite par une très forte hausse des prix sur le mécanisme de capacité pour les années de livraison 2020 et 2021 principalement (de l'ordre de 25 k€/MW). Cette évolution est la confirmation que l'architecture décentralisée et sa temporalité ont permis de renvoyer un signal-prix en lien avec les aléas affectant l'équilibre offre-demande, même à moins d'un an de l'échéance, ce qui constitue une particularité du mécanisme de capacité français.

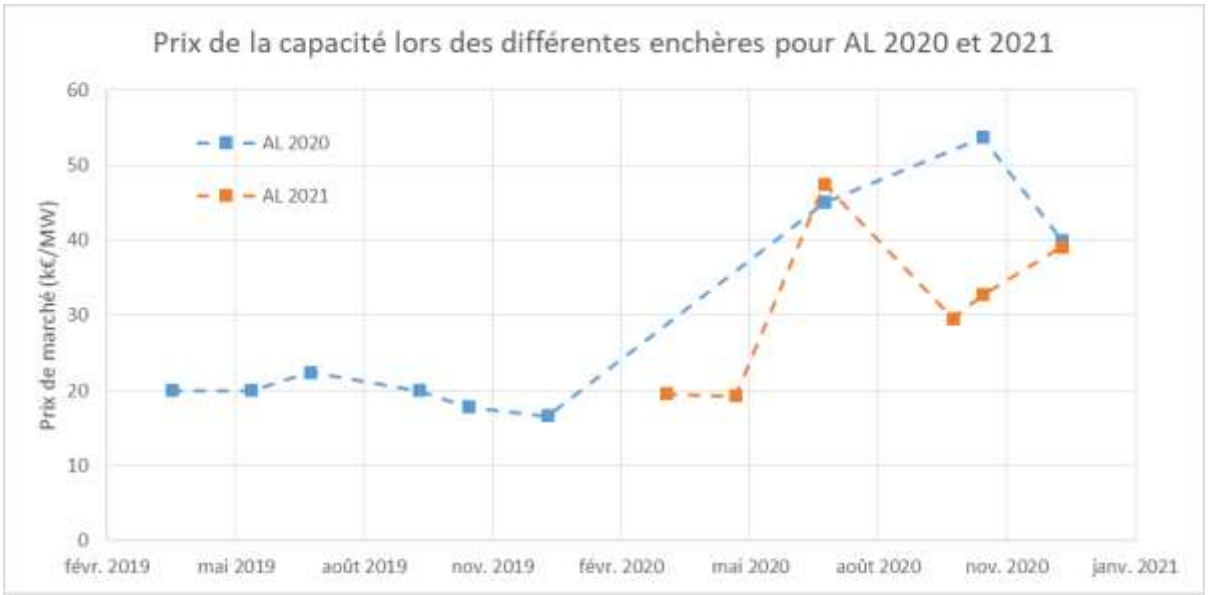


Figure 14 – Evolution des prix de la capacité suite à la crise sanitaire pour les années 2020 et 2021

[24] RTE, 2020, *Equilibre offre-demande pour l'hiver 2020-2021*, https://assets.rte-france.com/prod/public/2020-11/Rapport_hiver%202020-2021_novembre%202020%20DEF_0.pdf

[25] RTE, 2020, *Répercussions de la crise sanitaire sur l'approvisionnement en électricité pour l'hiver 2020-2021*, https://assets.rte-france.com/prod/public/2020-06/Analyse_preliminaire_hiver_2020-2021_-_VFinale-pdf.pdf

Le recours à un marché décentralisé ne réduit pas le risque d'une telle situation qui était par nature difficilement anticipable mais assure que le coût n'en soit pas transféré à la collectivité. En effet, les acteurs responsables de cette situation se trouvent en écart sur leur périmètre qui, sans rééquilibrage, conduira à des pénalités lors des règlements financiers, tandis que le prix de la capacité peut aider à l'émergence de capacités utiles à la sécurité d'approvisionnement (à un coût plus faible que celui d'une coupure d'électricité pour les consommateurs).

5.6.2. … ce qui a aidé à l'émergence de capacités améliorant la sécurité d'approvisionnement

Le prix de la capacité constitue en effet un signal économique renvoyé à tous les acteurs susceptibles de dégager des marges qui peut être mis au crédit de l'architecture décentralisée retenue : lorsque le prix monte, les consommateurs sont incités à réduire leur obligation et de nouvelles capacités utiles à la sécurité d'approvisionnement peuvent être mobilisées. La situation atypique vécue au cours de l'année 2020 et l'augmentation du prix de la capacité qui s'en ait suivi, a permis de mesurer ces effets[26].

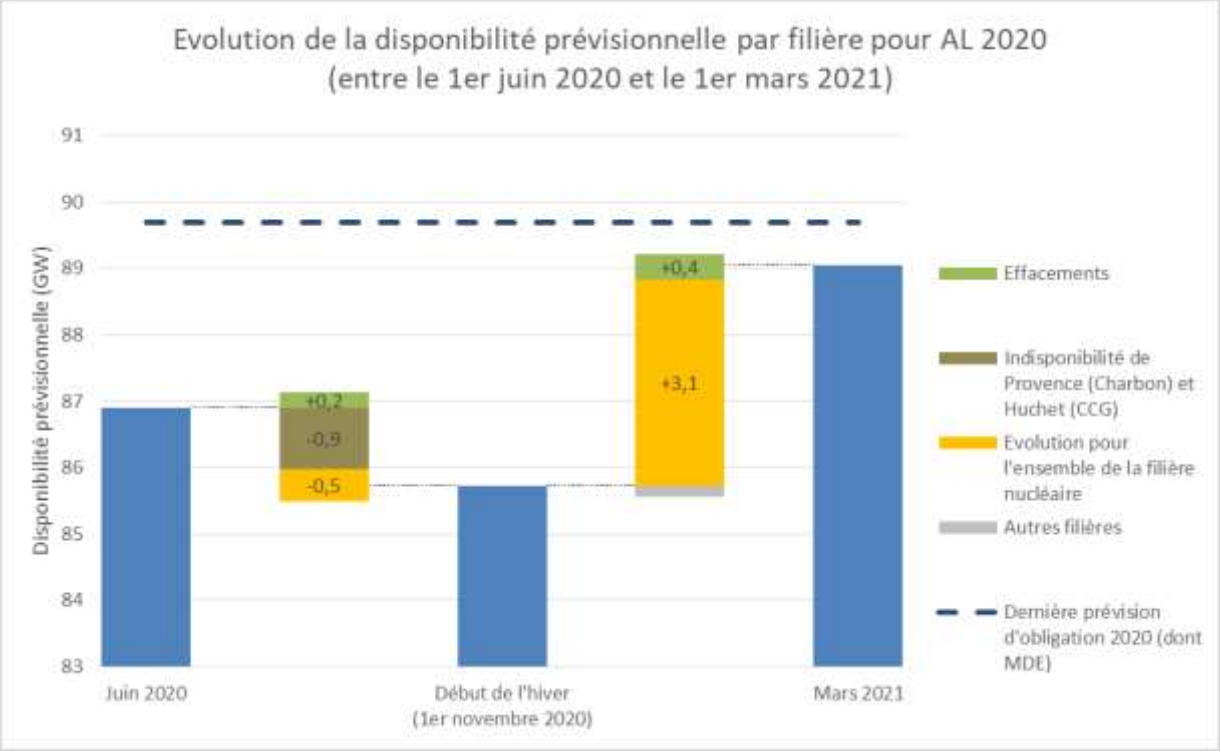


Figure 15 – Evolution de la disponibilité par filière pour l'année de livraison 2020 (entre juin 2020 et début mars 2021)

RTE estime que l'évolution des prix sur le mécanisme de capacité pour les années 2020 et 2021 a conduit à la mobilisation de capacité :

- + 600 MW d'effacement de consommation, qui peuvent possiblement être imputés à l'évolution des prix.
- + 2,5 GW d'évolution de la disponibilité du nucléaire et notamment le choix d'EDF d'ajuster ses plannings en réduisant la disponibilité du nucléaire sur l'été 2020 pour la maximiser sur

[26] Seule la réaction du côté de l'offre de capacité a été évaluée compte tenu des délais pour connaître avec précision le niveau de la consommation lors de l'hiver 2020-2021.

l'hiver 2020-2021. Cette évolution peut s'expliquer par l'enjeu de sécurité d'approvisionnement en électricité des français, qui s'est traduit par une élévation du prix de la capacité pour les années de livraison 2020 et 2021. Sans préjuger de l'impact financier de la crise sanitaire pour EDF, les revenus liés à l'amélioration de la disponibilité nucléaire[27] sur l'hiver 2020-2021 suite au réaménagement du planning sont pour 30% de type capacitaire compte tenu des niveaux de prix de la capacité observés pour AL 2020 depuis le début de la crise sanitaire et des prix spot réalisés sur la même période.



Figure 16 – Evolution du niveau de disponibilité globale pour AL 2020 et 2021 depuis le début de la crise sanitaire, comparé aux dernières prévisions d'obligation (septembre 2020)

Alors qu'un déficit de capacité était clairement établi pour AL 2020 de l'ordre de 4 GW, cette mobilisation a permis de rétablir l'équilibre entre l'offre et la demande de capacité.

Néanmoins, il est délicat d'établir le rôle exact du signal-prix de la capacité dans ces évolutions car d'autres facteurs explicatifs viennent s'y superposer :

- Une fois l'année de livraison débutée, une partie de l'évolution de la disponibilité des capacités peut être le résultat de la disponibilité constatée lors des jours PP2 tirés. Compte tenu de la meilleure disponibilité du parc nucléaire en décembre 2020 qu'en novembre, le placement des jours PP2 au mois de décembre[28] a conduit à une évaluation de la disponibilité nucléaire supérieure à celle anticipée (sur une prévision statistique de distribution des jours PP2 lors des mois de novembre et décembre 2020).
- La tension sur l'équilibre offre-demande à l'origine de l'augmentation du prix de la capacité se traduit également par une hausse des prix sur les marchés de l'énergie incitant les exploitants à maximiser leur production.

[27] Entre la disponibilité constatée et anticipée du nucléaire avant réaménagement du planning
[28] Hormis le 30 novembre

- Pour la filière effacement, une modalité exceptionnelle de l'appel d'offre effacement 2021 attribuait un bonus pour les capacités disponibles dès le 1er novembre 2020 et dans une moindre mesure l'évolution des règles du mécanisme de capacité ont permis d'annuler certains frais qui se seraient appliqués à des nouvelles capacités.

Par ailleurs, il convient d'évaluer l'émergence de nouvelles capacités (dont amélioration de la disponibilité) à l'aune de la durée séparant l'apparition du signal-prix de la période de livraison qui pour l'hiver 2020-21 s'est avérée particulièrement courte (6 mois) et de la volatilité du prix de la capacité qui rend incertain le revenu que peuvent espérer ces nouvelles capacités.