

NOTE METHODOLOGIQUE SUR LES INDICATEURS DE TENSION

En septembre 2020, la Dares et Pôle emploi ont rénové la grille d'analyse des tensions sur le marché du travail, en diffusant :

- d'une part, un **indicateur principal** qui permet de classer les métiers en fonction de leur degré de tension ;
- d'autre part, des indicateurs dits **complémentaires**, visant à comprendre le ou les facteurs à l'origine d'une éventuelle tension détectée sur un métier donné.

Cette fiche détaille le **mode d'élaboration de ces indicateurs annuels de tension**, mis à disposition sur le champ de la France hors Mayotte et sur les années 2011 à 2020. Elle précise tout d'abord comment est évalué l'indicateur principal de tension (cf. partie I), ainsi que les indicateurs complémentaires (cf. partie II). Elle revient ensuite sur les principes retenus pour la standardisation et la discrétisation de ces indicateurs (cf. partie III), avant de détailler les différentes sources mobilisées, ainsi que les niveaux de granularité retenus, à la fois sur le plan de la nomenclature des métiers que sur celui des échelons géographiques (cf. partie IV).

TABLE DES MATIERES

I. Indicateur principal de tension	3
II. Indicateurs complémentaires	4
II.1. Intensité d'embauches.....	4
II.2. Conditions de travail contraignantes	5
II.3. Non-durabilité de l'emploi.....	5
II.4. Main-d'œuvre disponible	6
II.5 Lien entre la spécialité de formation et le métier.....	7
II.5.1. Principe général.....	7
II.5.2. Calcul d'un indicateur synthétique de relation formation-emploi	7
II.6. <i>Mismatch</i> géographique	9
III. Sources et niveau de granularité pour appréhender les tensions sur le marché du travail.....	9
III.1. Sources de données	9
III.1.1. Statistiques du marché du travail.....	9
III.1.2. Offres collectées en ligne	9
III.1.3. Besoins en main-d'œuvre	10
III.1.4. Enquête Conditions de travail - Risques psycho-sociaux.....	10
III.1.5. Enquête Emploi en Continu	10
III.1.6. Enquêtes annuelles de recensement	10
III.2. Nomenclature des métiers	10
III.3. Territoires	11

I. Indicateur principal de tension

L'indicateur principal de tension est un indicateur composite. Il permet de fournir une échelle numérique de gradation des tensions par métier. Il se décline en trois composantes :

1. Le **rapport entre le flux d'offres d'emploi en ligne**, sur un champ étendu au-delà des seules offres collectées par Pôle emploi (cf. partie IV.1 pour une description plus détaillée de l'ensemble des sources mobilisées), **et le flux de demandeurs d'emploi inscrits à Pôle emploi en catégorie A** (sans emploi, tenus de rechercher activement un emploi).
2. Le **taux d'écoulement de la demande d'emploi**, qui mesure le taux de sortie des listes des demandeurs d'emploi de catégories A, B, C (sans emploi ou en activité réduite, tenus de rechercher activement un emploi).
3. La **part des projets de recrutements anticipés comme difficiles** par les employeurs.

Ces composantes représentent respectivement 30 %, 20 % et 50 % de l'indicateur synthétique. Ces pondérations sont établies à partir d'une analyse de données, en s'inspirant des contributions de ces variables aux axes tensions de l'analyse en composantes principales (cf. document d'étude à venir de la Dares sur la mesure des tensions).

L'indicateur de tension ($IT_{m,a,t}$) pour un métier donné m , l'année a et le territoire t est donc établi selon la formule suivante :

$$IT_{m,a,t} = 0.3 \times \left[\frac{T_{m,a,t} - \bar{T}}{\sigma_T} \right] + 0.2 \times \left[\frac{E_{m,a,t} - \bar{E}}{\sigma_E} \right] + 0.5 \times \left[\frac{D_{m,a,t} - \bar{D}}{\sigma_D} \right]$$

Où :

- $T_{m,a,t} = \frac{OEE_{m,a,t}}{DEE_{m,a,t}}$ est **rapport entre le flux d'offres d'emploi en ligne¹, et le flux de demandeurs d'emploi** inscrits à Pôle emploi en catégorie A (sans emploi, tenus de rechercher activement un emploi) pour le métier m , l'année a et le territoire t . Ce rapport correspond à l'ancien indicateur de tension (publié par la Dares et Pôle emploi jusqu'au second trimestre 2017) mais modifié pour couvrir un périmètre plus large que celui des seules offres collectées par Pôle emploi, puisqu'il prend également en compte des offres collectées en ligne par *scraping* par la Dares (cf. partie IV.1.2.).
 \bar{T} est la moyenne des ratios offres sur demandes d'emploi ($T_{m,a,t}$) tous métiers (regroupés en familles professionnelles²) et territoires confondus, calculée sur la période 2014-2018 et σ_T son écart type sur la même période.
- $E_{m,a,t} = \frac{DES_{m,a,t}}{DEFM_{m,a-1,t} + DEE_{m,a,t}}$ désigne le **taux d'écoulement de la demande d'emploi** pour le métier m , l'année a et le territoire t . Il est calculé comme le ratio entre :
 - le nombre de demandeurs d'emploi de catégories A, B, C (sans emploi ou en activité réduite, tenus de rechercher activement un emploi) sortis des listes de Pôle emploi durant l'année a ($DES_{m,a,t}$) ;
 - la somme de la moyenne sur les quatre fins de trimestre (mars, juin, septembre, décembre) de l'année précédente des demandeurs d'emploi de catégories A, B, C ($DEFM_{m,a-1,t}$) et du flux de demandeurs d'emploi entrés en catégories A, B, C durant l'année a ($DEE_{m,a,t}$). \bar{E} représente la moyenne des $E_{m,a,t}$ tous métiers (en FAP 87) et territoires confondus sur la période 2014-2018 et σ_E son écart type sur la même période.
- $D_{m,a,t} = \frac{(RD_{m,a,t} + RD_{m,a+1,t})/2}{(R_{m,a,t} + R_{m,a+1,t})/2}$ désigne la **part des projets de recrutements difficiles**. Elle est calculée à partir des réponses des employeurs dans l'enquête Besoins en main-d'œuvre de Pôle emploi (BMO) pour un métier donné m , l'année a et le territoire t . Elle est calculée comme le ratio entre :

¹ Ces offres d'emploi en ligne regroupent les offres collectées par Pôle emploi et les offres collectées en ligne par la Dares en utilisant le *scraping*.

² Ce niveau de nomenclature contient 87 catégories dont 77 sont couvertes par le champ tensions. Il est appelé FAP 87 par la suite.

- le nombre de **projets de recrutement anticipés comme difficiles** par les recruteurs ($RD_{m,a,t}$ et $RD_{m,a+1,t}$);
- le nombre total de **projets de recrutement anticipés** ($R_{m,a,t}$ et $R_{m,a+1,t}$).

L'enquête pour l'année a est réalisée entre octobre et décembre de l'année $a-1$ et les entreprises tendent à répondre à cette question à partir des difficultés qu'elles ont récemment rencontrées : la moyenne entre les projets attendus comme difficiles en a et en $a+1$ est donc considérée comme représentative des difficultés réellement rencontrées en a^3 .

\bar{D} est sa moyenne tous métiers (en FAP 87) et territoires confondus des $D_{m,a,t}$ sur la période 2014-2018 et σ_D son écart type sur la même période.

II. Indicateurs complémentaires

Les indicateurs complémentaires, au nombre de six, permettent d'éclairer les facteurs à l'origine des tensions, afin d'identifier la pluralité des situations selon les métiers et les territoires.

Tous les indicateurs présentés ici sont dans un premier temps standardisés, en considérant leur distribution en 87 familles professionnelles au niveau national sur la période 2014-2018 (période de référence), de façon à pouvoir les mettre en équivalence et les mobiliser conjointement. Dans un second temps, les indicateurs complémentaires sont discrétisés afin de faciliter l'analyse de la typologie des tensions. Les indicateurs complémentaires sont discrétisés en cinq classes (allant de 1 à 5), définies à partir des quintiles de leurs distributions en FAP 87 au niveau national sur la période 2014-2018.

II.1. Intensité d'embauches

Plus les employeurs recrutent, plus ils ont à rechercher des candidats et à réitérer le processus, ce qui joue potentiellement sur les tensions. Cette dimension est abordée en rapportant le nombre d'offres d'emploi et de projets de recrutement à l'emploi moyen. L'indicateur d'intensité d'embauche ($IE_{m,a,t}$) pour métier donné m , l'année a et le territoire t peut s'écrire comme suit :

$$IE_{m,a,t} = 0.5 \times \left[\frac{oeem_{a,t} - \overline{oee}}{\sigma_{oee}} \right] + 0.5 \times \left[\frac{r_{m,a,t} - \bar{r}}{\sigma_r} \right]$$

Où :

- $oeem_{a,t} = \frac{OEE_{m,a,t}}{N_{m,a,t}}$ désigne le **flux d'offres d'emploi en ligne pour 1 000 salariés** pour le métier m , l'année a et le territoire t . Il est calculé comme le rapport entre :
 - le nombre d'offres d'emploi ($OEE_{m,a,t}$) multiplié par 1 000 ;
 - le nombre de salarié du métier ($N_{m,a,t}$). \overline{oee} est la moyenne des flux d'offres d'emploi en ligne pour 1 000 salarié ($oeem_{a,t}$) tous métiers (en FAP 87) et territoires confondus, calculée sur la période 2014-2018 et σ_{oee} son écart type sur la même période.
- $r_{m,a,t} = \frac{(R_{m,a,t} + R_{m,a+1,t})/2}{N_{m,a,t}}$ désigne le **nombre de projets de recrutement pour 1 000 salariés** pour le métier m , l'année a et le territoire t . Il se calcule comme le rapport entre :
 - La demi-somme du nombre de projets de recrutement en a et en $a+1$ ($(R_{m,a,t} + R_{m,a+1,t})/2$) multiplié par 1 000 ;
 - le nombre moyen de salarié ($N_{m,a,t}$).

³ Cf. Note de révision de l'indicateur de tensions en 2020.

\bar{r} est la moyenne du nombre de projets de recrutement pour 1 000 salariés ($r_{m,a,t}$) tous métiers (en FAP 87) et territoires confondus, calculée sur la période 2014-2018 et σ_r son écart type sur la même période.

II.2. Conditions de travail contraignantes

Lorsqu'elles sont contraignantes, les conditions de travail peuvent accroître la difficulté qu'éprouvent les employeurs à attirer des candidats. Un indicateur synthétique sur les conditions de travail est évalué à partir de la part de salariés subissant des contraintes physiques, des limitations physiques, des contraintes de rythme, du travail répétitif, durant les jours non ouvrables ou en dehors des plages de travail habituelles et un morcellement des journées de travail. Pour le métier m et l'année a , il s'écrit de la façon suivante :

$$CT_{m,a} = \frac{1}{6} \sum_{i=1}^6 CT_{m,a,i}$$

Où :

- $CT_{m,a,1}$ est la **part des salariés du métier m qui, dans l'exécution de leur travail, font face à au moins trois contraintes physiques** parmi les suivantes : rester longtemps debout ; rester longtemps dans une autre posture pénible ou fatigante à la longue ; effectuer des déplacements à pied longs ou fréquents ; porter ou déplacer des charges lourdes ; effectuer des mouvements douloureux ou fatigants ; subir des secousses ou vibrations.
- $CT_{m,a,2}$ est la **part des salariés du métier qui sont limités** (fortement ou pas) depuis au moins six mois, à cause d'un problème de santé, dans les activités que les gens font habituellement.
- $CT_{m,a,3}$ est la **part des salariés du métier qui, dans l'exécution de leur travail, font face à au moins trois contraintes de rythme** parmi celles-ci : le déplacement automatique d'un produit ou d'une pièce ; la cadence automatique d'une machine ; d'autres contraintes techniques ; la dépendance immédiate vis-à-vis du travail d'un ou plusieurs collègues ; des normes de production ou des délais à respecter en une heure au plus ou des normes de production ou des délais à respecter en une journée au plus ; une demande extérieure (clients, public) obligeant à une réponse immédiate ou une demande extérieure (clients, public) n'obligeant pas à une réponse immédiate; les contrôles ou surveillances permanents (ou au moins quotidiens) exercés par la hiérarchie.
- $CT_{m,a,4}$ est la **part des salariés du métier dont le travail consiste à répéter** continuellement une même série de gestes ou d'opérations.
- $CT_{m,a,5}$ la **part des salariés du métier qui travaillent habituellement ou occasionnellement le samedi, le dimanche**, entre cinq heures et sept heures le matin, entre vingt heures et minuit et la nuit, entre minuit et cinq heures du matin.
- $CT_{m,a,6}$ la **part des salariés du métier dont la journée de travail est habituellement morcelée** en 2 périodes séparées par 3 heures ou plus.

Cet indicateur, comme les autres, est centré et réduit, donnant :

$$CTS_{m,a} = \left[\frac{CT_{m,a} - \overline{CT}}{\sigma_{CT}} \right]$$

Où \overline{CT} est la moyenne de l'indicateur sur les conditions de travail $CT_{m,a}$ tous métiers (en FAP 87) confondus sur la période 2014-2018 et σ_{CT} son écart type.

II.3. Non-durabilité de l'emploi

Le type d'emploi proposé peut avoir une incidence sur son attractivité et donc sur l'apparition de tension. Dans cette perspective, un indicateur synthétique est élaboré pour mesurer la « non-durabilité » des emplois : ainsi, plus cet indicateur est élevé, moins l'emploi est potentiellement attractif, plus les tensions sont accrues.

La non-durabilité de l'emploi est mesurée par l'opposé de la moyenne pondérée de la part des offres durables, de la part des offres à temps complet et de la part de projets de recrutement non saisonniers. Elle peut s'exprimer de la façon suivante pour un métier donné m , l'année a et le territoire t :

$$NDE_{m,a,t} = - \left\{ 0.5 \times \left[\frac{oeeed_{m,a,t} - \overline{oeeed}}{\sigma_{oeeed}} \right] + 0.25 \times \left[\frac{oeeec_{m,a,t} - \overline{oeeec}}{\sigma_{oeeec}} \right] + 0.25 \times \left[\frac{rs_{m,a,t} - \overline{rs}}{\sigma_{rs}} \right] \right\}$$

Où :

- $oeeed_{m,a,t} = \frac{OEED_{m,a,t}}{OEE_{m,a,t}}$ est la **part des offres d'emploi en CDI ou en CDD de plus de six mois** pour un métier donné m , l'année a et le territoire t . Il est calculé comme le rapport entre :
 - le nombre d'offres d'emploi en CDI ou en CDD de plus de six mois dans le métier m ($OEED_{m,a,t}$) ;
 - le nombre d'offres d'emploi dans le métier m ($OEE_{m,a,t}$). \overline{oeeed} est sa moyenne tous métiers et territoires confondus des $oeeed_{m,a,t}$ sur la période 2014-2018 et σ_{oeeed} son écart type sur la même période.
- $oeeec_{m,a,t} = \frac{OEEC_{m,a,t}}{OEE_{m,a,t}}$ est la **part des offres d'emploi à temps complet** pour un métier donné m , l'année a et le territoire t . Il est calculé comme le rapport entre :
 - le nombre d'offres d'emploi à temps complet dans le métier m ($OEEC_{m,a,t}$) ;
 - le nombre d'offres d'emploi dans le métier m ($OEE_{m,a,t}$). \overline{oeeec} est sa moyenne tous métiers et territoires confondus des $oeeec_{m,a,t}$ sur la période 2014-2018 et σ_{oeeec} son écart type sur la même période.
- $rs_{m,a,t} = \frac{(RS_{m,a,t} + RS_{m,a+1,t})/2}{(R_{m,a,t} + R_{m,a+1,t})/2}$ représente la **part des recrutements anticipés comme non saisonniers** par les recruteurs pour un métier donné m , l'année a et le territoire t . Il est calculé comme le rapport entre :
 - la demi-somme du nombre de projets de recrutement anticipés comme non saisonniers dans le métier m en a et en $a+1$ ($(RS_{m,a,t} + RS_{m,a+1,t})/2$) ;
 - la demi-somme du nombre de projets de recrutement en a et en $a+1$ ($(R_{m,a,t} + R_{m,a+1,t})/2$). \overline{rs} est sa moyenne tous métiers (en FAP 87) et territoires confondus des $rs_{m,a,t}$ sur la période 2014-2018 et σ_{rs} son écart type sur la même période.

II.4. Main-d'œuvre disponible

Une main-d'œuvre disponible trop réduite contribue à accroître les tensions. Afin de tenir compte de ce phénomène, un indicateur est élaboré en prenant l'opposé du nombre de demandeurs d'emploi en catégorie A, à l'emploi moyen. Il peut être formulé de la manière suivante pour un métier donné, l'année a et le territoire t :

$$MOD_{m,a,t} = - \left[\frac{defm_{m,a,t} - \overline{defm}}{\sigma_{defm}} \right]$$

où $defm_{m,a,t} = \frac{DEFM_{m,a,t}}{N_{m,a,t}}$ désigne le **nombre moyen de demandeur d'emploi en catégorie A sur les quatre fins de trimestre d'une année pour 1 000 salariés** dans le métier m , l'année a et le territoire t . Il est calculé comme le rapport entre :

- le nombre de demandeur d'emploi de catégorie A recherchant un emploi dans le métier m ($DEFM_{m,a,t}$) multiplié par 1 000 ;
- le niveau d'emploi salarié du métier m ($N_{m,a,t}$).

\overline{defm} est la moyenne des $defm_{m,a,t}$ tous métiers (en FAP 87) confondus sur la période 2014-2018 et σ_{defm} son écart type.

II.5 Lien entre la spécialité de formation et le métier

Un décalage entre les compétences requises par les employeurs et celles dont disposent les personnes en recherche d'emploi peut être l'une des causes des tensions. Pour approcher ce phénomène, un indicateur est élaboré de façon à identifier si le métier en question est difficile d'accès pour des personnes ne possédant pas la formation requise, à partir de la spécificité et de la concentration des spécialités de formation par métier.

II.5.1. Principe général

Le lien spécialité de formation-emploi⁴ traduit l'intensité de la relation entre le métier et les spécialités de formation dont dispose les personnes qui l'occupe. Lorsque les personnes exerçant le métier ont des spécialités de formation spécifiques ou très concentrées sur quelques-unes, ce lien est très fort (valeur ++) ou fort (+). Dans le cas contraire, le lien est faible (valeur -) ou très faible (valeur --). Il peut prendre également une valeur médiane (o). Ces signes résultent de la discrétisation des indicateurs en cinq classes à partir de leurs quintiles.

Ces valeurs sont calculées en utilisant de façon combinée trois indicateurs :

1. un **indicateur de concentration des spécialités de formation**, calculé par l'indicateur d'Herfindhal (somme des carrés des parts de chaque spécialité de formation au sein de chaque métier). Le champ est celui des salariés débutants dans le métier. Il peut prendre potentiellement des valeurs entre 0 (lorsque la concentration est minimale) et 1 (lorsque la concentration est maximale).
2. un **indicateur de spécificité des spécialités de formation**. Cet indicateur est la somme divisée par deux, des valeurs absolues des écarts entre le poids de chaque spécialité dans le métier (champs des salariés) et le poids de cette même spécialité de formation dans l'ensemble des actifs en emploi.
3. la **part des non diplômés dans le métier parmi les salariés jeunes actifs**. Cette valeur s'élève à 8 % en moyenne et varie d'un métier à l'autre.

II.5.2. Calcul d'un indicateur synthétique de relation formation-emploi

Les indicateurs sont calculés à partir des enquêtes Emploi de 2009 à 2018 parmi les jeunes actifs (personnes ayant terminé leur formation initiale depuis au plus 10 ans et n'ayant pas repris des études). Dans le calcul des indicateurs de concentration et de spécificité, sont exclues les personnes non diplômées (sans diplôme ou avec le seul brevet des collèges) ou avec une spécialité de formation non renseignée.

L'indicateur synthétique est ensuite calculé en trois temps. Un indicateur synthétique provisoire est tout d'abord obtenu à partir de la moyenne des indicateurs de concentration et de spécificité, à l'exception des cas d'opposition forte entre une valeur négative de l'indicateur de concentration et une valeur positive de l'indicateur de spécificité : dans ce cas, la spécificité est privilégiée pour l'indicateur synthétique. Toutes les combinaisons possibles de l'indicateur de concentration et de spécificité sont considérées (cf. tableau 1). Ainsi par exemple, lorsque le niveau de concentration est « + » et le niveau de spécificité « ++ », il en résulte l'attribution d'un niveau « ++ » à l'indicateur synthétique provisoire (cf. 1^{re} ligne du tableau 1).

⁴ Lainé, F. (2005). De la spécialité de formation au métier: cas du bâtiment, de l'hôtellerie-restauration-alimentation et du commerce. *Economie et Statistique*, 388(1), 145-169.

Tableau 1 - Construction de l'indicateur synthétique provisoire à partir de la concentration et de la spécificité des spécialités de formation

Concentration des spécialités de formation	Spécificité des spécialités de formation	Indicateur synthétique provisoire
+	++	++
+	+	+
+	0	+
+	-	0
+	--	-
0	++	++
0	+	+
0	0	0
0	-	-
0	--	--
-	++	+
-	+	+
-	0	0
-	-	-
-	--	-
--	++	+
--	+	-
--	0	--
--	-	-
--	--	--

Dans un second temps, l'indicateur synthétique définitif est obtenu en fonction de la part des non diplômés (cf. tableau 2). Ainsi par exemple, lorsque l'indicateur synthétique provisoire est « ++ », l'indicateur définitif (cf. 1^{re} ligne du tableau 2) :

- est « + » si la part des non diplômés du métier est comprise entre 1,6 et 2,5 fois la moyenne tous métiers confondus ;
- est « 0 » si la part des non diplômés est supérieure à 2,5 fois la moyenne.

Ainsi, lorsque la part des non diplômés est élevée, la valeur de l'indicateur synthétique est « rétrogradée » en lui affectant une valeur moins élevée : son calcul sur les seuls diplômés conduit en effet à surestimer le rôle des spécialités de formation dans les métiers où la part des non diplômés est élevée.

Tableau 2 - Construction de l'indicateur synthétique définitif en fonction de la part des non diplômés

Indicateur synthétique provisoire	Part des non diplômés comprise entre 1,6 et 2,5 fois la moyenne	Part des non diplômés supérieure à 2,5 fois la moyenne
++	+	0
+	0	-
0	-	--
-	-	--
--	--	--

Enfin, ces indicateurs sont transformés en valeurs numériques (de 1 à 5) pour la construction des indicateurs finaux.

II.6. Mismatch géographique

L'indicateur de *mismatch* géographique vise à rendre compte de la dimension spatiale des tensions sur le marché du travail. Il est défini comme l'écart, en valeur absolue, de distribution géographique entre l'offre et la demande d'emploi, selon l'indice de dissimilarité de Duncan et Duncan :

$$MG_{m,a} = \frac{1}{2} \sum_{ZE} \left| \frac{OEE_{m,a,ZE}}{OEE_{m,a}} - \frac{DEFM_{m,a,ZE}}{DEFM_{m,a}} \right|$$

Où :

- $OEE_{m,a,ZE}$, $OEE_{m,a}$ désigne le nombre d'offres d'emploi du métier m , l'année a , dans la zone d'emploi ZE et au niveau national pour l'année a , respectivement. Ainsi, le ratio entre les deux donne la part des offres du métier m durant l'année a dans la zone d'emploi ZE .
- $DEFM_{m,a,ZE}$, $DEFM_{m,a}$ désigne le nombre de demandeurs d'emploi de catégorie A du métier m dans la zone d'emploi ZE et au niveau national durant l'année a . Le ratio de ces deux grandeurs correspond à la part des demandeurs d'emploi du métier dans la zone d'emploi.

Cet indicateur a aussi été standardisé à l'instar des autres déjà présentés, en considérant sa distribution sur la période 2014-2018, bien qu'elle n'évolue pas durant la période. L'indicateur retenu *in fine* s'obtient par la formule suivante pour un métier donné m et l'année a :

$$MGS_{m,a} = \left[\frac{MG_{m,a} - \overline{MG}}{\sigma_{MG}} \right]$$

où \overline{MG} est la moyenne des $MG_{m,a}$ tous métiers (en FAP 87) confondus sur la période 2014-2018 et σ_{MG} est son écart type.

III. Sources et niveau de granularité pour appréhender les tensions sur le marché du travail

III.1. Sources de données

L'approche développée ici vise à explorer les différentes dimensions des tensions sur le marché du travail, en mobilisant différentes sources de données.

III.1.1. Statistiques du marché du travail

Les statistiques du marché du travail (STMT), issues des fichiers de gestion de Pôle emploi, décrivent les demandes et les offres d'emploi. Elles comptabilisent les demandeurs d'emploi en fin de mois ainsi que les flux d'entrées et de sorties des listes. Elles fournissent également les offres d'emploi déposées par les employeurs auprès de Pôle emploi.

III.1.2. Offres collectées en ligne

Pour compléter le champ des offres d'emploi déposées à Pôle emploi, la Dares collecte des offres diffusées en ligne. Les sites ciblés visent à couvrir tous les métiers et l'ensemble du territoire. Les offres sont collectées par *scraping* de l'ensemble des offres diffusées sur ces sites.

La classification de ces offres par métier est réalisée grâce à un algorithme de *machine learning* (*Support Vector Machine*, *SVM*) entraîné sur plus d'un million d'offres déposées à Pôle emploi annotées par code ROME (nomenclature du répertoire opérationnelle des métiers). Cet algorithme détermine le métier à partir du titre de l'offre d'emploi. Avant classification, les titres des offres sont normalisés en éliminant les *stopwords* (« à », « de », « une », etc.), explicitant les d'abréviations et en procédant à une lemmatisation.

Une fois les offres en ligne classées suivant leur code ROME, il convient de les basculer dans la nomenclature des familles professionnelles (FAP). Cette opération s'effectue traditionnellement en appliquant une matrice de passage nécessitant de disposer de la qualification du poste. Or, cette dernière n'est pas disponible sur tous les sites. C'est la

raison pour laquelle une autre table de passage permettant de basculer directement des ROME aux FAP est mobilisée. Cette table est obtenue à partir des croisements ROME x FAP observés sur le champ des offres déposées à Pôle emploi et des offres *scrapées* contenant la variable de qualification.

Enfin, afin de ne pas compter plusieurs fois une même offre diffusée sur plusieurs sites, les offres en ligne sont dé-dupliquées. Pour déterminer les offres en double, les offres disposant du même code ROME, dans le même département, correspondant au même nom d'entreprise et ayant été publiées dans un intervalle de deux semaines (14 jours) sont regroupées. Au sein de chaque groupe, les offres sont comparées deux à deux en calculant la similarité de Jaccard entre leurs descriptions (corps de l'offre). Si cette similarité est supérieure à 0,95, les offres sont considérées comme étant les mêmes ; alternativement, elles sont supposées distinctes.

III.1.3. Besoins en main-d'œuvre

Pôle emploi réalise annuellement une enquête sur les besoins en main-d'œuvre (BMO) des employeurs. Les questions qui leur sont posées portent, pour chaque métier, sur leurs projets de recrutement pour l'année à venir, en particulier pour ceux qu'ils anticipent comme difficiles et/ou saisonniers.

L'enquête est réalisée en septembre et porte sur les projets de recrutement de l'année suivante. L'observation des réponses montre qu'elle s'appuie sur des anticipations et les contrats sur le passé récent. Les indicateurs relatifs à l'année a (nombre de projets de recrutement, part des projets difficiles, nombre de projets de recrutements saisonniers) sont construits en calculant la moyenne des réponses pour l'année a (fournies entre septembre et décembre a-1) et pour l'année a+1 (fournies entre septembre et décembre a).

III.1.4. Enquête Conditions de travail - Risques psycho-sociaux

L'enquête conditions de travail – risques psycho-sociaux est conduite tous les trois ans par la Dares. Elle vise à obtenir une description concrète du travail, de son organisation et de ses conditions selon divers angles, ainsi que des risques associés. Ce sont les dernières éditions de 2013 et 2016 et 2019 qui sont mobilisées ici, les trois seules compatibles avec la période d'analyse des tensions sur le marché du travail (2011-2020).

III.1.5. Enquête Emploi en Continu

L'enquête Emploi en Continu est réalisée chaque trimestre par l'Insee. Elle est mobilisée ici pour fournir la structure de l'emploi par métier au niveau national (2015-2019) mais aussi pour analyser la relation entre formation et emploi par métier (2009-2018).

III.1.6. Enquêtes annuelles de recensement

Les enquêtes annuelles de recensement sont réalisées chaque année par l'Insee. Elles servent ici à décliner l'emploi par métier au niveau local.

III.2. Nomenclature des métiers

La nomenclature des familles professionnelles (FAP) de la Dares dans sa version de 2009 permet une analyse de l'emploi et du chômage par métier. Ces métiers sont regroupés en 225 familles professionnelles détaillées. L'analyse des tensions est menée à ce niveau de nomenclature. Certaines familles détaillées sont toutefois exclues, faute de couverture suffisante par les différentes sources mobilisées (cf. tableau 3).

Tableau 3 - Familles professionnelles détaillées exclues de l'analyse des tensions

FAP 225	Libellé FAP 225	Raison d'exclusion	FAP 87 filtrés
A0Z00	Agriculteurs indépendants	Non-salariés	
A0Z01	Éleveurs indépendants	Non-salariés	
A0Z02	Bûcherons, sylviculteurs indépendants	Non-salariés	
A1Z00	Maraîchers, horticulteurs indépendants	Non-salariés	
A1Z01	Viticulteurs, arboriculteurs indépendants	Non-salariés	
F3Z40	Artisans du travail du bois et de l'ameublement	Non-salariés – pas couvert par BMO	
K0Z20	Ouvriers non qualifiés divers de type artisanal	Non-salariés	X
K0Z40	Artisans et ouvriers qualifiés divers de type artisanal	Non-salariés	X
L6Z00	Dirigeants de petites et moyennes entreprises	Non-salariés	X
L6Z90	Cadres dirigeants des grandes entreprises	Non-salariés	X
P0Z60	Agents des impôts et des douanes	Fonction publique	X
P0Z61	Employés des services au public	Fonction publique	X
P0Z62	Employés de la Poste et des télécommunications	Fonction publique	X
P1Z80	Contrôleurs des impôts et des douanes	Fonction publique	X
P1Z81	Autres cadres B de la fonction publique	Fonction publique	X
P1Z82	Professions intermédiaires de la Poste et des télécommunications	Fonction publique	X
P2Z90	Cadres A de la fonction publique (hors spécialités juridiques) et assimilés	Fonction publique	X
P2Z91	Cadres de la Poste et des télécommunications	Fonction publique	X
P2Z92	Cadres de l'armée et de la gendarmerie	Fonction publique	X
P3Z90	Professionnels du droit	Fonction publique	X
P3Z91	Magistrats	Fonction publique	X
P4Z60	Agents de sécurité et de l'ordre public	Fonction publique	X
P4Z61	Agents de polices municipales	Fonction publique	X
P4Z80	Cadres intermédiaires de la police et de l'armée	Fonction publique	X
R1Z66	Vendeurs généralistes	Codage à partir du ROME insatisfaisant	
R3Z81	Intermédiaires du commerce	Codage à partir du ROME insatisfaisant	
S3Z00	Patrons d'hôtels, cafés, restaurants	Non-salariés	
V2Z90	Médecins	Non-salariés	X
V2Z91	Dentistes	Non-salariés	X
V2Z92	Vétérinaires	Non-salariés	X
V2Z93	Pharmaciens	Non-salariés	X
V5Z00	Exploitants d'équipements sportifs et culturels	Codage à partir du ROME insatisfaisant	
W0Z80	Professeurs des écoles	Fonction publique	X
W0Z90	Professeurs du secondaire	Fonction publique	X
W0Z91	Directeurs d'établissement scolaire et inspecteurs	Fonction publique	X
W0Z92	Professeurs du supérieur	Fonction publique	X
X0Z00	Professionnels de la politique	Catégorie difficile à saisir	X
X0Z01	Clergé	Catégorie difficile à saisir. Marché du travail spécifique.	X

Note de lecture : les trois FAP 225 que sont A0Z00, A0Z01, A0Z02 sont exclues de l'analyse car les sources mobilisées ne permettent pas de couvrir les non-salariés ; la FAP 87 A0Z peut toutefois être analysée, sur le champ des salariés. Les trois FAP 225 que sont P0Z60, P0Z61 et P0Z62 sont exclues de l'analyse car les sources mobilisées ne permettent pas de couvrir la fonction publique ; dans ce contexte, la FAP 87 P0Z ne peut pas être analysée.

III.3. Territoires

L'analyse des tensions sur le marché du travail est déclinée sur l'ensemble de la France (hors Mayotte), ainsi que sur l'ensemble des régions et départements, lorsque la couverture des données est suffisante. L'analyse est considérée comme possible sur un territoire donné si au moins 30 offres sont déposées dans l'année, 30 projets de recrutement

sont formulés par les entreprises et 30 demandeurs d'emploi sont inscrits sur les listes en catégorie A. Ce seuil est fixé à 10 pour la Corse et ainsi que pour les départements et régions d'Outre-mer.

Lorsque les volumes ne sont pas suffisants, le niveau de nomenclature métier ou d'échelon géographique immédiatement supérieur est proposé dans l'analyse. Plus précisément, lorsqu'un croisement famille professionnelle détaillée x niveau géographique ne peut pas être fourni faute de volumétrie suffisante, il est proposé :

- De renvoyer à la famille professionnelle à laquelle appartient le métier concerné, sur l'échelon géographique en question ;
- Si l'option précédente n'est pas non plus possible, de renvoyer à la famille professionnelle détaillée concernée mais sur une zone géographique plus large qui inclut le territoire considéré.
- Si l'option précédente n'est pas non plus possible, de réitérer le processus.