EBC
Energy in Buildings and Communities Programme

IEA EBC Annex 94:

Validation and verification of in-situ building energy performance measurement techniques

Working Meeting #2

May 11-12, 2026 - Paris, France

Contents

1. Participants

52 Participants, from 13 countries, joined this second working meeting.

Country	Name	Organisation	
Austria	Yousef Al Aidy	TU Wien	In Person
Austria	Philip Zettl	AEE INTEC	Online
Austria	Michael Gumhalter	AEE INTEC	Online
Belgium	Hans Bloem	DYNASTEE	In Person
Belgium	Alexander Ulrix	KULEUVEN	In Person
Belgium	Staf Roels	KU Leuven	In Person
Belgium	Martin Prignon	Buildwise	In Person
Belgium	Liesje Van Gelder	BCCA	In Person
Canada	Tyler Willson	SAIT	Online
Denmark	Kevin Smith	DTU Civil & Mechanical Engineering	In Person
Denmark	Peder Bacher	DTU	In Person
Denmark	Kim Wittchen	Aalborg University, Dept. of the Built Environment	In Person
Denmark	Matthias Van Hove	Ghent University and DTU	In Person
Denmark	Johannes Nicolás Wildfeuer	Technical University of Denmark (DTU)	In Person
France	Myriam Humbert	Cerema	In Person
France	Asmae Moutaoikil	University of Bordeaux I2M	In Person
France	Julien Waeytens	Université Gsutave Eiffel	In Person
France	guillaume Perrin	Université Gustave Eiffel	In Person
France	Sarah Juricic	Centre Scientifique et Technique du Bâtiment	In Person
France	Dave Ron	CEEP - Centre Energie Environnement Procédés	Online
France	Patrick Schalbart	Mines Paris, CEEP	In Person
France	Baptiste Poirier	Cerema	In Person
France	Simon Rouchier	Université Savoie Mont Blanc	In Person
France	Mirado Ravelonarivo	CSTB & UGE	In Person
Germany	Johannes Pernpeintner	Deutsches Zentrum für Luft- und Raumfahrt e.V. DLR	In Person
Germany	Ouiam Bounab	University of Abdelmalek Essaadi	Online
Ireland	Aimee Byrne	Trinity College Dublin	In Person
Italy	Riccardo Gazzin	Eurac Research	Online
Italy	Giuseppe De Michele	Eurac Research	Online
Netherlands	Rob Meester	Saxion, University of Applied Sciences	In Person
Netherlands	Twan Rovers	Saxion University of Applied Sciences	In Person
New Zealand	Griffin Cherrill	Building Research Association of New Zealand (BRANZ)	Online
Spain	Pablo Hernández	University of the Basque Country (UPV/EHU)	In Person
Spain	Aitor Erkoreka	University of the Basque Country (UPV/EHU)	In Person
Spain	Jose Alberto Díaz Angulo	CIEMAT	Online

Spain	Luis Yélamos Martín	Universidad de Almería - CIEMAT	Online
Spain	Beñat Arregi	Tecnalia	Online
Spain	Iñigo Lopez	Tecnalia Research & Innovation	Online
Spain	Juan Maria Hidalgo	University of the Basque Country UPV/EHU	Online
Sweden	Justinas Smertinas	Technical University of Denmark	In Person
UK	Mark Collett	Leeds Beckett	In Person
UK	Sam Stamp	UCL IEDE	In Person
UK	David Glew	Leeds Beckett University	In Person
UK	Richard Fitton	UoS	In Person
UK	Kevin Gornall	UK Department for Energy Security & Net Zero	In Person
UK	Virginia Gori	UCL	In Person
UK	Cliff Elwell	UCL Energy Institute	In Person
UK	Steven Heath	Knauf Insulation	Online
UK	Vicente Orts Mercadillo	Vector Labs	In Person
UK	Carlos Jimenez-Bescos	University of Westminster	In Person
UK	Lior Carno	Kestrix	In Person
UK	Kami Rakhshanbabanari	Loughborough University	Online
UK	Max Eastwood	BRE	Online
UK	Richard Jack	Build Test Solutions	In Person
UK	David Johnston	Leeds Beckett University	Online
UK	Matt Goodridge	Kestrix	In Person
UK	Matthew Li	Loughborough University	In Person
UK	Jamie Corson	UCL	In Person
UK	David Allinson	Loughborough University	In Person
UK	Grant Henshaw	University of Salford	In Person

2. Agenda

Monday May 11, 2026

11.00 Arrival and lunch
12.00 Welcome and Keynote speaker
12:30 Annex 94 update
13.00 Sub Task 1: progress update, survey findings, and plans
14:00 Sub Tasks 2&3: workshop on variability and seasonality of the HTC
15:00 Presentations of work relevant to the annex
16:30 Dynastee conference planning
17.00 Visit to Sense City

Tuesday May 12, 2026

08:30 Arrival and coffee
09.00 Aim of today and thoughts on publishing
09.30 Breakouts for Sub Tasks 2, 3, and 4
11:30 Feedback from breakouts
12.30 Sub Tasks 2&4: testing the methods
13.00 Lunch and photographs
14.00 Sub Task 4: diagnostic tools
15.00 Sub Task 5: Data collection and curation
16:00 Thanks, details of the next meeting, and finish at 16:30

3. Day 1 May 11, 2026

Welcome and keynote speaker

Sarah Juricic welcomed us to CSTB.

Thomas Berthou gave a keynote presentation on the French EPC system and class-change risk. The analysis was based on five archetypes, representative of the French housing stock, before and after retrofit. Sobol analysis was used to understand the sensitivity of continuous variables, and on-at-a-time for discrete. Important values were floor area, heating system efficiency, system installation period, and construction period (used to determine U-value). Monte Carlo uncertainty analysis generated class-change risk where initial position in EPC band had no impact = 90% confidence interval / band width. Risk was low for renovated dwellings but tended to be high for non-renovated dwellings. Recommendations included integrating construction data from land registry tax database into EPC software, using the EPREL performance data from Ecodesign (requires a translator) and prohibiting the use of default values for recent HVAC systems.

Annex 94 update

Richard Fitton gave an overview of the Annex 94 objectives, sub tasks and deliverables. Progress is good to date. Next meeting will be in Madrid in late October alongside a DYNASTEE symposium. First peer reviewed academic journal paper from the annex has been published.

Sub Task 1: progress update, survey findings, and plans

Liesje Van Gelder and Mark Collett reminded us about sub-task 1 objectives related to *applications and their requirements*. During the first project year, Subtask 1 focused on mapping related projects, developing shared terminology, and conducting a stakeholder survey to identify how measured HTC and energy performance data are used across sectors. The Paris meeting confirmed the list of mapped projects and had a final push for some more survey responses with the survey closing on the 20th May. Analysis and write up of the survey responses is planned for 2026 with an initial focus of the drivers and barriers to adopting HTC measurement. Outputs include white papers, impact report, development of an advisory board, and journal publications.

- D1.1 Mapping related projects
  - spreadsheet created – contact Liesje with any updates.
- D1.2 Non-technical vocabulary
  - work in progress with ST3.
- D1.3 Stakeholder survey
  - “How do we measure up” launched at the Vienna meeting last October.
  - 274 responses out of a target of 170.
  - Survey will be closed next week.
  - Responses are being translated into English.
  - Data will be analysed and a paper published.

Survey interim findings:

- Responses from 20 countries
- Mainly consultants, academic, government bodies, building performance bodies, etc…

- Most respondents understand HTC in principle, but many do not use in their work.
- Most not measuring HTC.
- Biggest challenges with measurement included uncertainty, client understanding, cost, time and disruption, lack of regulation.

Next steps include carrying out in depth interview, meeting, or workshops to understand survey response in more detail. This will lead to an impact report – what is needed to translate the work of this annex into real world applications. An advisory panel will be formed to support this. A specification for the advisory panel will be developed to aid recruitment to that panel.

Sub Tasks 2&3: workshop on variability and seasonality of the HTC

Katia Ritoša, María José Jiménez, Sarah Juricic, Frances Hollick, Staf Roels. and Alexander Ulrix:

- Uncertainty, variability and seasonality
- Reasons that the HTC varies and the need to not confuse variability with uncertainty
- Finding some variability is normal but when do you need to acknowledge variability
- Repeatability conditions vs intermediate conditions vs reproducibility conditions
- How this varies with building type and how many tests on how many types are required?
- Staf Roels presented a paper:
    - Can we apply methods on a wider scale?
    - Case study of 65 homes in the DEFACTO dataset: energy signature, averaging, single linear regression, Siviour method, multiple linear regression, autoregressive method.
    - What is the optimal length of the monitoring period? Seasonal average value?
    - Most reliable results from winter periods with high delta T and low solar.
    - ARX models worked better than regression in summer.
- Alexander Ulrix presented a paper:
    - Hypothesis that there will be a correlation between HTC and windspeed
    - 30-day moving windows applied to TEST project data – 21 out of 30 dwellings
    - Results show strong correlation with windspeed.
- Maria Jose presented a paper:
    - Sampling theorem – sampling frequency should be twice the period of variation in the model.
    - Impact of signal processing on dynamic models – can be a problem when it destroys information.
    - Need models that work between dynamic and steady state assumptions to suit frequency of input data?

Presentations of work relevant to the annex

Participants presented past research work relevant to Annex 94. Details can be seen on the Annex 94 website: https://annex94.iea-ebc.org/event?EventID=9727

#	Presenter	Company/Institution	Title	
1	Aitor Erkoreka Gonzalez	University of the Basque Country-	Uncertainty sources of the average method to estimate the HTC	In-person
2	Aimee Byrne	University of Dublin	An introduction to relevant Irish studies and related national priorities	In-person

3	Grant Henshaw	University of Salford	A study of the effect of snow on roofs and energy performance	In person
4	Hassam Rehman	VTT	Annex 93 (resilience in cold climates) and Annex 94	On-line
5	Asmae Moutaoikil	Universite Bordeaux	Adaptive facades (?)	In person

Dynastee Conference planning

Hans Bloem, Maria Jose, Richard Fitton, Aimee Byrne, Twan Rover

- Summer school
  - 14-18 September 2026
  - University of Almeria, Spain
  - Focus on whole building HTC assessment
  - Using new data this year – 50% lecture, 50% data analysis and interaction
- Newsletter
  - Invitation to sub-task leads to supply content by end of month
  - Next issue June 2026
- Symposium
  - Next meeting 27-29 October 2026 in Madrid at CIEMAT
  - First day is the symposium - focus on HTC measurement
  - Aim to bring together academia, industry, and government organisations
  - 1st session Annex 94 – 90 minutes – aimed at non-academic audience
  - Industry and government entities with Academic contributions
  - 0930 to 1300, 1400 to 1730
  - Social event and dinner in the evening (own expense)
  - 28-29 September is Annex 94 meeting
- Request for content for the Dynastee website from sub-task leaders

Visit to Sense City

4. Day 2 May 12, 2026

Aim of today

Julien Waeytens welcomed us to University Gustave Eiffel. Cliff Elwell remined us of the objectives of subtasks 2, 3, 4, and 5 and the work needed to drive matters forward. The importance of definitions and variability was emphasised and agreed to discuss further with feedback from the ST3 breakout.

Breakouts for Sub Tasks 2, 3, and 4

The **Subtask 2** breakout session continued the discussion on key challenges related to variability and uncertainty, concluding on the need for clearer definitions and mapping out sources of each, in connection with ST3. A central theme was how different methodological approaches and underlying assumptions can be aligned, especially regarding their dependence on preprocessing steps, which currently introduce inconsistencies across results. While CE3 demonstrates that harmonised preprocessing can yield robust and replicable outputs, many approaches remain tightly coupled to their own preprocessing choices, underscoring the need to systematically map simple preprocessing options and assess their impact (ongoing work by Alexander, current findings/problem statement reported by Pablo). Further, Justinas' contribution emphasised the relationship between white-, grey-, and black-box models and the associated trade-offs in information interpretation and loss, including efforts to quantify these effects. Building on initial CE3 results, the group plans to extend analyses to dynamic estimation methods (led by Simon and others). The work is progressing toward a new common exercise and datasets (expected June 2026), including the ConstrucThor digital twin, while continuing to address temporal aspects as well as boundary conditions (i.e., neighbouring buildings and ground effects).

The **Subtask 3** breakout session began with quick updates on the progress in the uncertainty and validation working groups, and then focussed on the ground truth task. An HTC definition dependent on the system boundary, conditions, observation period, and model method was presented by Matthias Van Hove, which led to a lively discussion in the group. Members of ST3 appreciated the simplicity of the definition, highlighting that this indicates it's been well thought through. There were discussions of potential confusion with the 'observation period' terminology and what that represents, and also some thoughts about what this means for the future of HTC measurement. It was decided to present this again to the wider Annex. We then presented the idea for a common exercise, exploring the opportunity to use energy consumption data as a ground truth - this has been proposed for some methods, and would be a great if it were able to be applied more widely, so exploring to what extent and in what scenarios good energy prediction => good HTC measurement is highly valuable as a task. Initially this will be an exercise for in-use methods, but we explored obtaining some test conditions data to explore within those methods also.

The **Subtask 4** breakout session focused on two main workstreams: the development of the diagnostic chatbot tool and the design of a field trial programme. The chatbot, built on a lightweight open-source model and inspired by WebMD's symptom-checker approach, would guide non-expert users through the ST4 diagnostic flowchart; the group agreed it is a valid annex output provided the method is documented and the code published openly, with a companion paper describing the system design rather than the code itself. On field trials, two case studies were presented, a 1970s mid-terrace and a new build with a >100% performance gap, to illustrate the

core diagnostic concept: disaggregating a measured HTC into fabric components (U-values, airtightness, ventilation) to identify where heat loss exceeds expectation, with any residual attributed to thermal bridging or measurement uncertainty. The group identified several existing data sets (Leeds Beckett TSBP/DEEP, EHL buildings, Saxion Smart TinyLab) as a starting point, agreed on a minimum measurement protocol, and noted that a mix of existing and new-build properties is needed. Additional tasks agreed included compiling an international table of building regs/overheating/underheating standards, and the planning of field trials, aimed at autumn 2026.

Feedback from breakouts

ST3. Matthias Van Hove presented a framework to define the HTC: *What is the true value of H? - A conditional measurand framework for H*. Sarah Juricic explained that further clarifications and examples would be developed and shared with the group. Next steps to consider how this translates into a validation framework for testing methods e.g. setting domains for comparison. Also, the validation paper is nicely on the way, seminar planned for July, herringbone paper will start soon. A new common exercise will be launched in the next few weeks to test validation by energy estimation.

ST2. Katia Ritoša presented results of a common exercise on the TEST project dataset using steady state methods. Results matched when data were pre-processed in the same way. Therefore, next steps to explore this preprocessing further and how it affects the results. Also, expanding to dynamic methods, moving to new data, and using simulated data from the ConstrucThor digital twin.

ST4. Richard Jack and Grant Henshaw presented their work on diagnosing a performance gap – carrying out a range of measurements to identify different problems. They are developing this work through a planned field trial for ~20 case study buildings. How can we collect more diagnostic data in future studies? Request for data sets where HTC was measured and additional tests carried out to enable breaking down the components of the gaps. Volunteers were identified.

Sub Tasks 2&4: testing the methods

Grant Henshaw explained the opportunity to collect new data during the next winter to collect data for use by the annex that would suit the aims of all the sub-tasks. One of Aimee Byrne's projects has developed templates for data sharing agreements that we could use. Richard Fitton asked who was collecting data next winter. Cliff Elwell highlighted the need to use opportunities from funded projects and setting up suitable data sharing agreements.

Sub Task 4: diagnostic tools

Vicente Orts Mercadillo presented an LLM that can support diagnosing energy performance. It uses data from the open EPC database and asks questions to lead the user through a diagnosis. People volunteered to provide EPC from each country to Richard Jack. Plan is to develop and test in a field trial.

ST5: Data collection and curation

Matt Li introduced the objectives of ST5 and reviewed progress against each activity and deliverable. Ten datasets have been recorded and attendees indicated that other datasets can be

logged – Matt will send a link round. Review of best practice data collection and management is ongoing – plan to leverage the work done by Aimee Byrne, and Annex 81. Future meeting to review consent forms and any changes that may be needed so that they work for the annex. Discussion around action plans for new data collection. Vienna meeting identified that existing datasets were not suitable. New experiments in test facilities, field trials, and simulations are required. Attendees met in sub task groups to complete a spreadsheet. The results were discussed as a group. Several attendees have test house that they can provide data from. Matt will generate a short summary for each sub-task.

Thanks, details of the next meeting and finish at 16:30

Sarah Juricic provided details of the evening meal and the Stakeholder Symposium that would take place tomorrow at Ecole des Mines de Paris: *From Practice to Standard*.

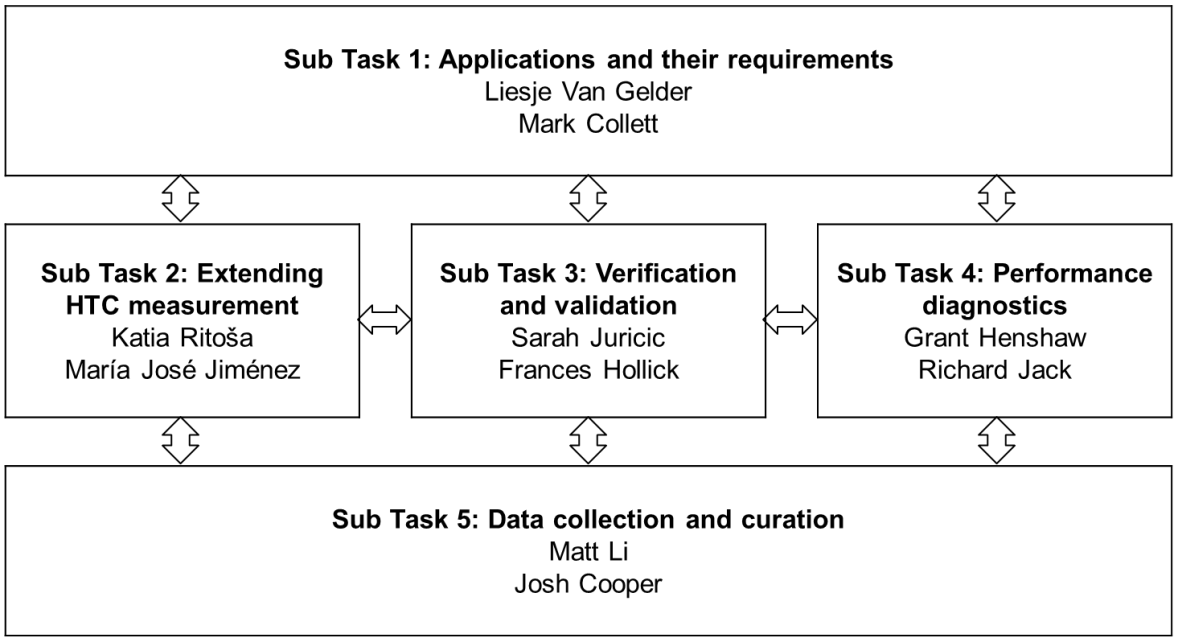
María José Jiménez provided details of the next meeting – 27-29 October at CIEMAT in Madrid.

David Allinson explained that we were at the end of the first of our three working-years and that this was the second meeting of six. He reflected on progress and described the annex 94 pathways to impact and what was needed to generate the outputs, to create the outcomes and thereby deliver that impact.

Problem	Input	Output	Outcome	Impact
Applications and their requirements are poorly understood.	ST1: collaboration between stakeholders.	New taxonomy and applications guide.	Confidence in methods grows.	Growth in measurement industry.
Methods may not work in all building types.	ST2: extend methods through collaborative R&D.	Proven methods for measurement, validation and verification.	New applications identified	
Methods mainly designed for use in winter.			New methods developed.	Feedback from measurement drives better performance.
Methods may be inaccurate.	ST3: develop and evaluate verification and validation.		Validation and verification standardised.	
Diagnosing a poor HTC result can be difficult.	ST4: advance diagnostics through collaborative R&D.			Performance gap reduced.
Lack of data sets for research & development	ST5: collect and curate data sets.	Opensource data sets.	Measurement applied at scale.	

The three operating agents thanked everyone for joining the meeting, in-person and on-line. They thanked the sub-task leaders for making progress. They thanked our excellent hosts at CSTB and University Gustave Eiffel: Sarah Juricic and Mirado Ravelonarivo; Julien Waeytens.

Meeting closed



General Enquiries		Annex94@salford.ac.uk
(Co) Operating Agents		
David Allinson	Loughborough University	D.Allinson@lboro.ac.uk
Richard Fitton	University of Salford	R.Fitton@salford.ac.uk
Cliff Elwell	UCL	clifford.elwell@ucl.ac.uk
Sub Task leaders		
Liesje Van Gelder	BCCA / SECO	l.van_gelder@bcca.be
Mark Collett	Leeds Beckett University	M.Collett@leedsbeckett.ac.uk
Katia Ritoša	K U Leuven	katia.ritosa@kuleuven.be
Maria José Jiménez	CIEMAT	mjose.jimenez@psa.es
Sarah Juricic	CSTB	Sarah.JURICIC@cstb.fr
Frances Hollick	UCL	frances.hollick.15@ucl.ac.uk
Grant Henshaw	University of Salford	g.p.henshaw@salford.ac.uk
Richard Jack	BTS	richard.jack@buildtestsolutions.com
Matt Li	Loughborough University	M.G.J.Li@lboro.ac.uk
Josh Cooper	Hildebrand	joshua.cooper@hildebrand.co.uk